
THE UNIVERSITY
*of* EDINBURGH

# INVESTIGATION OF THE CHROMATIN COMPOSITION AND STRUCTURE OF FOREIGN DNA IN A MAMMALIAN CELL

Maximilian Hamilton Fitz-James


Thesis submitted for the degree of Doctor of Philosophy

University of Edinburgh

2018

## Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where otherwise stated by reference or acknowledgment, the work presented is entirely my own.

Maximilian Hamilton Fitz-James
May 2018

## Abstract

In order to contain many millions, or even billions of base pairs within every nucleus of a eukaryotic cell, DNA must be extensively packaged. This is achieved by association of DNA with packaging proteins, resulting in the formation of chromatin, which can lead to various degrees of compaction. The most extreme form of compaction is the highly condensed mitotic chromosome, formation of which is necessary for proper resolution and segregation of the genetic material during cell division. However, the exact nature of the structure of chromatin within the mitotic chromosome and the factors which regulate it remain subjects of debate and continued investigation.

The hybrid cell line F1.1 presents a unique tool for the study of mitotic chromosome structure. This mouse cell line has been observed to present a distinct chromatin structure in mitosis assembled over a large region of DNA inserted into one of its chromosomes and originating from the fission yeast *Schizosaccharomyces pombe*. Direct comparison of the structure of this distinct region of chromatin with that of the adjacent endogenous chromatin could provide insight into the nature of mitotic chromosome structure as well as the properties of the chromatin which are influencing this structure. Microscopy and Hi-C analyses showed that the mitotic chromatin organising or “scaffold” proteins are not altered over the region of *S. pombe* chromatin, but that the amount of chromatin organised around these proteins is diminished. In accordance with the “radial-loop” model of mitotic chromosome structure, we put forward a model whereby the *S. pombe* chromatin is organised into smaller chromatin loops around a constant organising scaffold. Examination of the histone post-translational modifications over the region of *S. pombe* chromatin revealed it to be highly heterochromatic, with high levels of H3K9me3 and associated factors such as HP1α and 5meC, and low levels of activating marks. Generation of further mammalian – *S. pombe* fusion cell lines recapitulated both the distinct mitotic structure and the heterochromatic profile of the inserted *S. pombe* chromatin. However, insertion of *S. pombe* DNA into a mouse cell by transfection rather than fusion resulted in a large region of *S. pombe* DNA that lacked both a distinct structure and heterochromatin. These results suggest that H3K9me3-mediated heterochromatin may influence the structure of chromatin in mitosis, leading to an organisation into smaller chromatin loops than non-heterochromatic regions.

## Lay Summary

When cells divide one of their most important goals is to split the DNA evenly between the two new cells, so that each cell has the exact same DNA. To facilitate this splitting, DNA is tightly packaged during cell division. However, the exact way in which the DNA is condensed from its loose configuration to its compact configuration in cell division is not clear. When a large amount of DNA from a yeast species is inserted into a mouse cell, the inserted DNA has a different appearance to mouse DNA in cell division, suggesting that it is packaged differently. Identifying the differences between the yeast DNA and mouse DNA in this hybrid cell line is a useful way of discovering new factors that contribute to the packaging of DNA during cell division. By comparing the yeast and mouse DNA we form a model to explain their different appearances based on the DNA being packaged into loops of different sizes. We also found a particular chemical modification, called H3K9me3, on the proteins associated with the yeast DNA. By creating new hybrid cell lines by different methods, we show that those which have the H3K9me3 modification associated with the yeast DNA also have a similar distinct structure. On the other hand, hybrid cells which do not have H3K9me3 associated with the yeast DNA do not show a distinct structure, and appear the same as mouse DNA. This suggests that H3K9me3 is somehow affecting the structure of DNA during cell division.

## Table of Contents

## List of Figures

## Abbreviations

| | |
|---|---|
| 3C | Chromosome conformation capture |
| 5meC | 5' cytosine methylation |
| AGE | Agarose gel electrophoresis |
| AMP | Adenosine monophosphate |
| ATP | Adenosine triphosphate |
| bp | Base pairs |
| BPTF | Bromodomain PHD finger transcription factor |
| BSA | Bovine serum albumin |
| CAF-1 | Chromatin assembly factor 1 |
| cAMP | Cyclic AMP |
| CD | Chromodomain |
| CDK | Cyclin dependent kinase |
| CENP-A | Centromere protein A |
| ChIP | Chromatin immunoprecipitation |
| CpG | Cytosine-guanine dinucleotide |
| CREB | cAMP response element binding protein |
| CTCF | CCCTC-binding factor |
| CTD | Carboxy-terminal domain |
| dH2O | Distilled water |
| DMEM | Dulbecco's modified eagle's medium |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| DNMT | DNA methyltransferase |
| dNTP | Deoxynucleoside triphosphate (nucleotide) |
| dox | Doxycycline |
| D-PBS | Dulbecco's PBS |
| dsRNA | Double stranded RNA |
| DTT | 1,4-dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |

| | |
|---|---|
| EM | Electron microscopy |
| ESC | Embryonic stem cell |
| EYFP | Enhanced yellow fluorescent protein |
| FACT | Facilitates chromatin transcription |
| FBS | Fetal bovine serum |
| FISH | Fluorescence in situ hybridisation |
| FLC | Flowering locus C |
| GFP | Green fluorescent protein |
| GMEM | Glasgow's minimal essential medium |
| HAT | Histone acetyltransferase |
| HDAC | Histone deacetylase |
| HDM | Histone demethylase |
| HMT | Histone methyltransferase |
| Hox | Homeobox |
| HP1 | Heterochromatin protein 1 |
| IP | Input |
| ISWI | Immitation SWItch |
| JmjC | Jumonji C |
| JMJD2D | Jumonji domain containing protein 2D |
| KAP-1 | KRAB associated protein 1 |
| kb | Kilobase |
| KCl | Potassium chloride |
| KDM | Histone lysine demethylase |
| Kif4 | Kinesis family member 4 |
| KMT | Histone lysine methyltransferase |
| KRAB | Kruppel associated box domain |
| LAD | Lamin-associated domain |
| LB | Lysogeny broth |
| LIF | Leukemia inhibitory factor |
| lncRNA | Long non-coding RNA |
| MBD | Methyl-binding domain |

| | |
|---|---|
| mRNA | Messenger RNA |
| NaCl | Sodium chloride |
| NEAA | Non-essential amino acids |
| NEB | Nuclear enveloppe breakdown |
| NOR | Nucleolar organising regions |
| N-terminal | Amino-terminal |
| NuRD | Nucleosome remodelling deacetylase |
| NURF | Nucleosome remodelling factor |
| Oct4 | Octamer binding protein 4 |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| PFA | Paraformaldehyde |
| piRNA | Piwi-interacting RNA |
| PRC | Polycomb repressive complex |
| PRE | Polycomb response element |
| PTM | Post-translational modification |
| qPCR | quantitative PCR |
| RCA | Regulator of chromosome architecture |
| RNA | Ribonucleic acid |
| RT-qPCR | Reverse transcriptase |
| RT-qPCR | Reverse transcription qPCR |
| SAR | Scaffold attachment region |
| SDS | Sodium dodecyl sulphate |
| SEM | Standard error from the mean |
| SET | Su(var)3-9, enhancer of zeste and trithorax |
| SETDB1 | SET domain bifurcated 1 |
| SF-DMEM | Serum-free DMEM |
| siRNA | Small interfering RNA |
| SMC | Structural maintenance of chromosomes |
| SUV39H1/2 | Suppressor of variegation homolog 1/2 |

| | |
|---|---|
| SWI/SNF | SWItch / Sucrose non-fermentable |
| TAD | Topological associated domain |
| TE | Tris-EDTA |
| TopoIIα | Topoisomerase IIα |
| TRIM28 | Tripartite motif containing 28 |
| Tween | Polyoxyethylenesorbitan monolaurate |
| Wapl | Wings apart like homolog |
| YES | Yeast extract supplemented |
| ZFP | Zinc finger protein |

# Chapter 1: Introduction

## 1.1 Chromatin

### 1.1.1 Definition and Function

In eukaryotes DNA is organised within the nucleus into a nucleoprotein complex called chromatin. The fundamental unit of chromatin is the nucleosome, with the so-called “core particle” corresponding to 146 base pairs of DNA wrapped around a protein complex composed of histone proteins (Kornberg and Lorch 1999, Luger *et al.* 1997). Each nucleosome contains 2 each of the 4 core histones H2A, H2B, H3 and H4 which associate first as 2 heterodimers of H2A/H2B and a tetramer of H3/H4, and then further into an octamer around which the DNA is wrapped (Krude 1995). These nucleosome core particles assemble along the DNA in a “beads-on-a-string” conformation, with each core particle separated by a varying length of linker DNA which may be associated with and protected by the linker histone H1, depending on the organism, cell type or region in question (Happel and Doenecke 2009). Thus, the repeating unit of chromatin corresponds to around 200 base pairs of DNA, including linker, associated necessarily with 2 copies each of the 4 core histones and potentially one copy of the linker histone H1.

This basic organisation is the first level of compaction required to fit approximately 2 metres of DNA in some eukaryotic cells into a nucleus with a diameter of a few micrometres (Woodcock and Gosh 2011). However, chromatin also serves many other functions such as protecting DNA from degradation as well as providing a regulatory role. While the presence of nucleosomes poses something of an obstacle for the accessibility of DNA, and is thus a generally repressive entity, various chemical modifications of the N-terminal tails of histones, which extend out from the mainly globular domains in the centre of nucleosomes, can inform the cellular machinery further and lead to activation or repression of gene expression by recruiting other factors or altering the interaction between nucleosomes (Campos and Reinberg 2009). This coupled with the chromatin remodelling activity of several proteins and complexes, which allows for a remarkable plasticity of chromatin, results in a highly dynamic structure that is permissive and informative to cellular

processes such as transcription, DNA replication and DNA repair (Chen and Dent 2014, Bell *et al.* 2011, Groth *et al.* 2007). These functions will be discussed in more detail below.

### 1.1.2 Chromatin assembly and remodelling

#### 1.1.2.1 Histone Chaperones

Assembly of histones onto DNA is performed by a number of histone chaperone proteins, which help carry out all the necessary functions of transport, nuclear import and nucleosome assembly and disassembly (de Koning *et al.* 2007). After DNA replication, nucleosomes are assembled downstream of the replication fork by replication-coupled chaperones such as the CAF-1 complex, which is also associated with nucleosome reassembly after DNA repair, while other chaperones are coupled to other cellular processes such as FACT with transcription. Chaperones such as HIRA also deposit or exchange histones independently of replication and transcription, assuring replacement or filling of any gaps where necessary. Some chaperones are associated with variant histones, which are generally similar to one of the four core histones with differences in protein sequence varying from minor to substantial and which play important roles in regulating or marking regions of DNA for specific functions. The most well-known of these is the H3 variant CENP-A, which is assembled at centromeres and forms the basis for kinetochore assembly in most eukaryotes (Henikoff and Smith 2015). Other variants include H2AZ, which often marks the first nucleosome of eukaryotic genes, and H2AX, deposited in its pre-phosphorylated form γH2AX to mark regions of DNA damage for repair (Henikoff and Smith 2015, Bell *et al.* 2011, de Koning *et al.* 2007).

#### 1.1.2.2 Chromatin Remodellers

Once assembled onto DNA, the position of the nucleosomes can be altered by the action of ATP-dependent chromatin remodellers. These remodellers are grouped into four families: SWI/SNF, ISWI, INO80 and NURD/Mi-2/CHD (Clapier and Cairns 2009). These proteins act by sliding or evicting nucleosomes for a variety of purposes. Certain members of the ISWI family, for instance, are associated with the nucleosome assembly machinery and are responsible for the regular spacing of

nucleosomes along the chromatin fibre and thus particularly important for the reorganisation of chromatin after replication (Haushalter and Kadonaga 2003, Saha *et al.* 2006). The SWI/SNF family on the other hand uses its sliding ability to make the normally occluded DNA accessible to transcription factors and the transcriptional machinery and is thus important in gene expression and regulation (Saha *et al.* 2006, Martens and Winston 2003).

### 1.1.3 Chromatin Marks

#### 1.1.3.1 Histone post-translational modifications

Histones are subject to a number of post-translational modifications (PTMs), primarily but not exclusively on their N-terminal tails, which extend out from the nucleosome core and are thus easily accessible. A wide variety of these modifications have been and continue to be described (Kouzarides 2007, Bannister and Kouzarides 2011) (**Figure 1.1 A**). Among the more common and best studied of these histone PTMs are acetylation of lysine residues, phosphorylation of serine or tyrosine residues and, perhaps the most varied, methylation of lysine or arginine residues, which in the former case can be mono-, di- or tr-methylation and in the latter can be mono-, symmetrical di- or asymmetrical di-methylation.

These histone PTMs are mediated by a number of chromatin modifying enzymes, many of which are well-characterised (Carlberg and Molnár 2016). In general, the levels of a given histone PTM will be determined by the antagonistic activity of two opposing classes of modifying enzyme. In the case of histone acetylation a large number of histone acetyl-transferases (HATs) are responsible for their deposition, while histone deacetylases (HDACs) catalyse their removal. Likewise, histone lysine methyl-transferases (KMTs) and demethylases (KDMs) are responsible for the deposition and removal of methyl groups, respectively. These enzymes can be classed into a number of families based on homology, but in general they will contain a chromatin binding domain such as a bromodomain, chromodomain or PHD finger, contributing to their specific recruitment to their target sites. In addition to this they will possess a catalytic domain specific to their modification. Thus, almost all KMTs, with the exception of Dot1, possess a SET domain that catalyses their specific methylation reaction, while many KDMs possess a JmjC domain

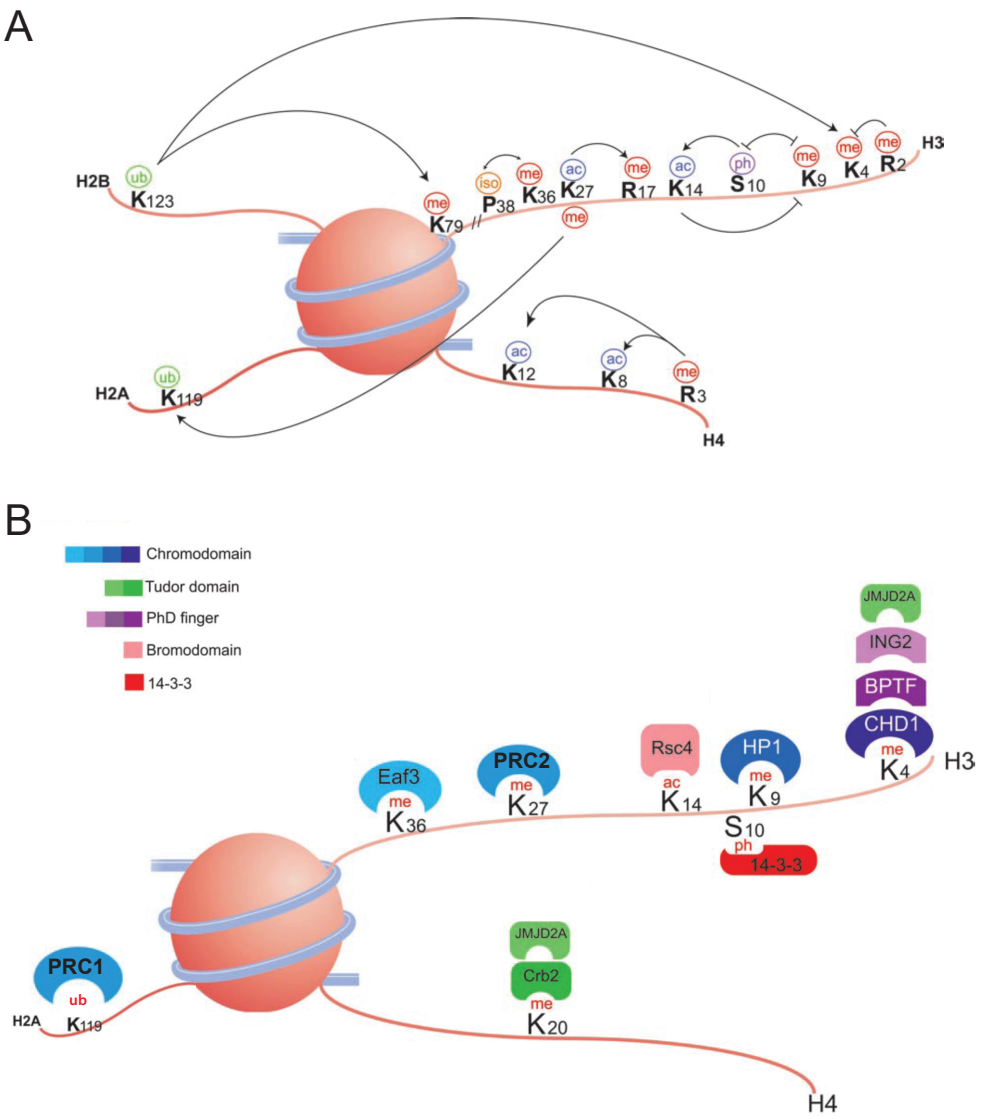


**Figure 1.1 - Overview of some well-described histone modifications and their binding proteins**

**A** - Diagram of several histone modifications including methylation, acetylation, phosphorylation and ubiquitination and their positive or negative association with each other.
**B** - Overview of some of the main proteins known to bind modified residues of histone tails sorted by the nature of their binding domain
Adapted from Bannister and Kouzarides 2011.

catalysing the demethylation of specific lysine residues (Qian and Zhou 2006, Dimitrova *et al.* 2015, Klose *et al.* 2006).

Different roles and mechanisms of action have been demonstrated or hypothesized for various PTMs, but they are primarily known for their effect on gene expression, for which the mechanism is thought to fall into one of two categories: direct influence on the organisation of chromatin by affecting the interaction between histones and DNA or between histones and other histones, and indirect effects through regulation of the binding of effector proteins. Due to the presence of high numbers of lysine residues the histone tails carry a considerable positive charge, which is thought to facilitate interaction with the negatively-charge backbone of DNA. Addition of negatively-charged acetyl or phosphate groups thus has the potential to neutralise this charge, particularly acetylation which can occur in large numbers at the many lysine residues along the tail (Bannister and Kouzarides 2011, Zhang *et al.* 1998, Hong *et al.* 1993). These modifications are thus thought to lead to a general opening-up of the chromatin fibre leading to greater accessibility. This idea is supported by observations of an increase in DNase sensitivity of acetylated regions (Simpson 1978) and some specific observations of acetylation marks preventing compaction *in vitro* (Shogren-Knaak *et al.* 2006).

Recruitment of effector proteins is a far more varied role for histone PTMs and can extend to those modifications which do not alter the charge of the proteins, including methylation. A number of PTMs have been found to lead to recruitment of transcriptional activators or repressors, chromatin remodellers, elements of the transcriptional machinery, or indeed other chromatin modifiers, leading to various associations between PTMs and the possibility of positive or negative feedback loops (Bannister and Kouzarides 2011) (**Figure 1.1 B**). An equally diverse role also lies in the opposite ability of PTMs to block binding of effector proteins, such as has been observed with the active gene mark H3K4me3 preventing the binding of the repressive NuRD complex (Zegerman et al. 2002).

Histone PTMs thus form an essential part of the chromatin environment which informs the cellular machinery and can lead to profound changes in the organisation of the chromatin as well as the expression state of the underlying genetic sequence. The pattern of histone PTMs has led to the idea of a “histone code” by which the

state or function of a gene or region of chromatin can be inferred by interpretation of the chromatin marks it bears, though the term “code” is perhaps too strict a term, suggesting rigorous adherence rather than what is in effect a tendency (Jenuwein and Allis 2001). Histone PTMs have thus become a major component of that information which has been given the label “epigenetic”, i.e. “above the genetic information” (though the term has been attached to different definitions).

#### 1.1.3.2 DNA methylation

DNA itself can also be chemically modified directly. The most well-known type of such modification is the methylation of cytosine bases on the 5’ position of the pyrimidine ring (5meC), which occurs in vertebrates at CG dinucleotides or CpGs (Day and Sweatt 2010, Bird 2002, Bogdanovic and Veestra 2009) (**Figure 1.2**). This methylation is initially catalysed by the *de novo* DNA methyltransferases (DNMTs) DNMT3a and DNMT3b, which are similar in sequence and have overlapping though not entirely redundant roles, as both are essential for viability in mice (Okano *et al.* 2002). These *de novo* DNMTs are most active at two points in development that see large-scale demethylation of DNA, followed by establishment of new patterns of methylation: first during gametogenesis and then again during pre-implantation development (Morgan *et al.* 2005). The methylated state of CpGs is then maintained through cell division by the maintenance methyltransferase DNMT1, which acts by recognising hemi-methylated CpGs on double-stranded DNA and methylating the unmethylated strand (Jeltsch 2006).

60-90% of vertebrate CpGs are methylated throughout the genome, the major exception being clusters of highly concentrated unmethylated CpGs called CpG islands (Bird 1987). These CpG islands coincide with gene promoter regions and are important reference points for the transcriptional machinery. In general, it can thus be said that methylation of CpGs is the norm and its absence is informative. Indeed, certain DNA-binding proteins bind unmethylated CpGs, thus targeting them specifically to CpG islands and, by extension, gene promoters, a key example being the polycomb associated H3K4 and H3K36 lysine demethylase KDM2B (Farcas *et al.* 2012). On the other hand, methylated DNA also exerts a strong repressive effect on a region of DNA by several mechanisms. Firstly, methylated cytosine can directly interfere with the binding of sequence-specific transcription factors, likely through

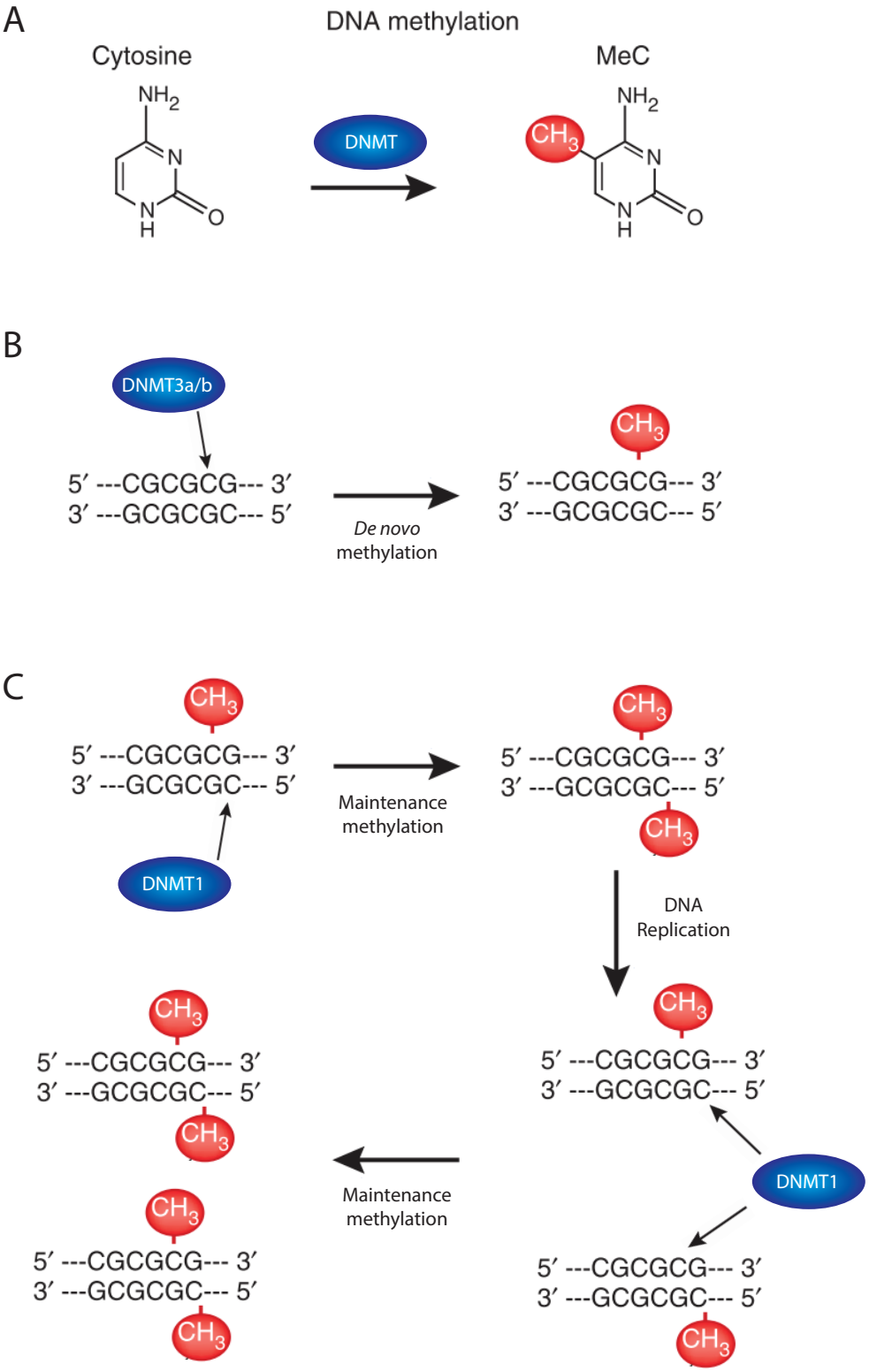


**Figure 1.2 - Principles of DNA methylation establishment and maintenance**

Chemical basis of DNA methylation on cytosines (**A**) and processes of DNA methylation at CpG dinucleotides by de novo methyltransferases DNMT3a/b (**B**) and maintenance methyltransferase DNMT3 (**C**).
Adapted from Day and Sweatt 2010.

steric effects, as has been observed with E2F and CREB (Campanero *et al.* 2000, Iguchi-Ariga and Schaffner 1989). Secondly, Methylated CpGs are bound preferentially by a group of proteins in the MBD (methyl-binding domain) family, including MeCP2, MBD1, MBD2, MBD3 and MBD4 (though the last two have a more complicated binding pattern) (Fatemi *et al.* 2006). These proteins in turn are known to recruit a wide variety of other proteins, including histone modifiers such as HDACs, chromatin remodellers and transcriptional repressors (Bogdanovic and Veestra 2009). A major role of these repressive functions is the differential silencing of genes inherited from different parents in a process known as imprinting (Ferguson-Smith 2011), and DNA methylation is also a key factor in mammalian X chromosome inactivation (Heard *et al.* 2004).

#### 1.1.3.3 Euchromatin and Heterochromatin

Historically chromatin has been divided, initially based on observation of the nucleus by microscopy, into open, accessible and thus active “euchromatin” and condensed, inaccessible and inactive “heterochromatin” (Passarge 1979, Jenuwein and Allis 2001). It is now known that classification into one of these two categories is too simplistic a model and often inappropriate, however certain aspects of this conception have endured and may be useful.

The term euchromatin now generally denotes certain chromatin profiles characteristic of transcriptionally active regions, and thus a generally open chromatin state (Smolle and Workman 2013). Among the marks that distinguish these active states are a general hyperacetylation of histones, leading to a more open configuration of the chromatin as described above. H3K4me is also present, with promoter regions often tri-methylated and gene bodies di-methylated at H3K4, and is thought to recruit, among other proteins, chromatin remodelling factors such as the NURF complex through its subunit BPTF (Wysocka et al. 2006). H3K4 methylation is catalysed by the Set1/MLL family of HMTs, whose function has been linked in several organisms to the prior presence of H2B ubiquitylation (Eissenberg and Shilatifard 2010). H3K36me and H3K79me are also found at active genes in many species (Smolle and Workman 2013). This last PTM, catalysed by the Dot1 protein, is notable in that it occurs within the globular domain of the histone rather than the tail. It has been proposed that it acts to destabilise the nucleosome and

open up the chromatin in a mechanism not unlike that of histone acetylation but using steric effects rather than electrostatic ones (Lu *et al.* 2008). Its role is not fully understood however.

Heterochromatin on the other hand can be further divided into “facultative” and “constitutive” heterochromatin. Facultative heterochromatin is widespread in both plants and animals and is distinguished by the presence of H3K27me3 and H2AK119ub. These marks are deposited by the polycomb complexes PRC2 and PRC1, respectively (Margueron and Reinberg 2011, Hennig and Derkacheva 2009, Gan *et al.* 2015). The relationship between these complexes, their functions and the mechanism of their recruitment are areas of ongoing investigation (Blackledge *et al.* 2015). In *Drosophila melanogaster*, where PRC complexes were first discovered, recruitment occurs by a relatively straightforward sequence-based recognition of Polycomb Response Elements (PREs) (Steffen and Ringrose 2014). In vertebrates and plants the situation is more complicated, and recruitment appears to occur by a combination of various factors including DNA-binding factors, long non-coding RNAs (lncRNAs) and interaction with other chromatin marks (Blackledge *et al.* 2015, Brockdorff 2013, van Kruijsbergen *et al.* 2015, Hennig and Derkacheva 2009, Bemer and Grossniklaus 2012). In flowering plants, the recruitment of PRC2 to the FLC and HoxD loci by the lncRNAs COLDAIR and HOTAIR, respectively, have become key model pathways in understanding its recruitment mechanism (Kim and Sung 2012). In mammals on the other hand, much recent research has focused on the role of KDM2B in recruiting PRC1 through its recognition of unmethylated CpGs (Farcas *et al.* 2012), as well as the different roles of variant PRC complexes (differing in one or a few subunits) in recruitment and chromatin modification (Blackledge *et al.* 2014). Crucially, independent of their histone modifying activity, both PRC1 and PRC2 variants have been shown to be able to compact chromatin both *in vitro* and *in vivo* (Francis *et al.* 2004, Eskeland *et al.* 2010, Margueron *et al.* 2008), a role which may be linked to the polymerisation of one of the subunits of PRC1 (Isono *et al.* 2013).

Facultative heterochromatin is usually localised on small gene-rich domains, notably serving to silence key developmental regulators. In *Arabidopsis*, dynamic H3K27 methylation and demethylation is a key determinant of cell fate specification and the continuous differentiation of new organs by tight regulation of various transcription factors (Bemer and Grossniklaus 2012, Gan *et al.* 2015). In metazoa, the large

family of Homeobox (Hox) transcription factors are among the most important targets of PRC2 (Kim and Kim 2012). Larger domains of H3K27me3 are also possible however, as X-chromosome inactivation in mammals proceeds by recruitment of PRC2 through the long non-coding RNA Xist, by which H3K27me3 is deposited broadly across the chromosome (Heard *et al.* 2014). For the most part however, facultative heterochromatin is believed to be a more temporary or regulatory silencing mechanism acquired and potentially lost relatively quickly during development (Leeb and Wutz 2012). Its role should be considered together with the antagonistic activating H3K4me3 mark. In animals, the balance struck between these two marks can lead to so-called “bivalent” domains of chromatin that can be considered “poised” between the more permanent activated or silenced states that they will acquire later in development (Bernstein *et al.* 2006).

Constitutive heterochromatin is characterised by H3K9me3, a histone PTM which in mammals can be deposited by any of the KMTs SUV39H1, SUV39H2, SETDB1, G9A or GLP (Kim and Kim 2012). With the exception of budding yeast, in which heterochromatin formation is mediated by the Sir family of proteins (Hickman *et al.* 2012), A primary role of H3K9me3 is achieved by its recruitment of the HP1 proteins (Swi6 in *S. pombe*), whose mammalian homologues are HP1α, HP1β and HP1γ, the last of which is however also associated with some euchromatic regions (Lomberk *et al.* 2006). HP1 is recruited to constitutive heterochromatin by binding of H3K9me3 through its chromodomain (Bannister *et al.* 2001, Lachner *et al.* 2001). In the first instance HP1 leads to the strengthening and spreading of the heterochromatic marks through recruitment of SUV39 proteins in a positive feedback loop (Allshire and Madhani 2017). In mammals HP1 also recruits the Suv4-20h HMTs and several DNMTs and HDACs, leading to heterochromatic H4K20me3, DNA methylation and histone de-acetylation, respectively (Schotta *et al.* 2004, Fuks *et al.* 2003, Zhang *et al.* 2002). In a second instance, HP1 is believed to lead to compaction of chromatin (Danzer and Wallrath 2004), which may be a result both of the recruitment of other chromatin associated proteins and remodellers (Martens and Winston 2003, Lomberk *et al.* 2006) and by the action of HP1 itself, possibly linked to its ability to dimerise through its chromoshadow domain (Canzio *et al.* 2011).

Constitutive heterochromatin is primarily found at centromeric and telomeric regions, which are both composed of large arrays of tandem repeat sequences, and is a

ubiquitous feature of these domains (Kim and Kim 2012). In the former case H3K9me3 has been found to be essential for proper centromere function (Bernard *et al.* 2001, Kellum and Alberts 1995, Shimura *et al.* 2011, Molina *et al.* 2016). It also helps ensure genome integrity by preventing recombination at repetitive regions (Grewal and Jia 2007). H3K9me-mediated silencing has also been reported on smaller, unique sequences, and the diversity of H3K9 HMTs may reflect the PTM's different roles in silencing. While SUV39H1 and SUV39H2 are responsible for H3K9me3 deposition at centromeres and telomeres, G9A, GLP and SETDB1 have been found to be responsible for silencing at a number of euchromatic loci (Peters *et al.* 2001, Tachibana *et al.* 2002, Rice *et al.* 2003, Shinkai and Tachibana 2011). Finally, H3K9me3 is known to be the primary means for cells to silence parasitic genetic elements, including long interspersed DNA elements and long terminal repeats (LTRs) (Mikkelsen et al. 2007, Timms *et al.* 2016, Cuellar *et al.* 2017).

The mechanism by which constitutive heterochromatin is established and maintained therefore has an added layer of complexity due to its different targets and functions. Various DNA-binding proteins have been found to interact with mammalian H3K9 HMTs and have been proposed as contributing to their targeting to specific sites of DNA. SETDB1 has been linked to a well-known silencing pathway through its binding to KAP-1/TRIM28, which mediates its interaction with the large number of proteins in the Krüppel-associated box domain zinc-finger family (KRAB-ZFPs) that contribute to its targeting (Schultz *et al.* 2002). Oct4 has also been found to interact with SETDB1 and has been proposed to assist in silencing certain target genes in ES cells (Yeap *et al.* 2009, Yuan *et al.* 2009). The YY1 protein is another possible H3K9me3 recruiter to repetitive loci as it coincides with H3K9me3 at a number of repetitive elements and YY1 knock-down results in the depletion of H3K9me3 (Kim and Kim 2012, Wu *et al.* 2009).

A more general mechanism, particular in relation to repetitive sequences, may lie in the recruitment of H3K9me3 by small RNAs and their associated machinery (Holoch and Moazed 2015). In *S. pombe* dsRNAs of 22-23 nucleotides transcribed from pericentromeric repeats are cleaved by Dcr1 to form small interfering RNAs (siRNAs) which results in recruitment of the H3K9 KMT Clr4 to the centromeric region by the RITS complex and thus heterochromatin establishment (Allshire and Ekwall 2015). A similar process has been found in *Arabidopsis* in which the siRNA

pathway recruits DNA methylation to particular target sites, which in turn results in H3K9 methylation (Matzke and Mosher 2014). While RNAi-mediated silencing is not conserved in this form in metazoa, the related Piwi-interacting RNAs (piRNAs) have been shown to silence transposons in the germline of both *Drosophila* and mouse (Malone and Hannon 2009). piRNAs originate from single-stranded RNA precursors and are amplified by a “ping-pong” cycle of cleavage of complementary RNAs. H3K9me-dependent silencing of most transposons in the germ line have been found to be dependent on the piRNA-interacting protein Piwi, which may result from initial DNA methylation, similar to *Arabidopsis*, or directly from H3K9 KMT or HP1 targeting (Holoch and Moazed 2015).

The situation is complicated further in higher eukaryotes by the association of H3K9me3 and DNA methylation (Rose and Klose 2014). As mentioned previously, the presence of H3K9me3 can lead to recruitment of DNMTs via HP1 in mammals. However, the directionality or independence of the relationship between the two marks is highly complex. For instance, the MBD family proteins MeCP2 and MBD1 have been shown to interact with SUV39H1/H2 and SETDB1, respectively (Fuks *et al.* 2003, Sarraf *et al.* 2004), but these recruitment mechanisms are not essential for heterochromatin maintenance and thus likely redundant (Jiang *et al.* 2011, Zhao *et al.* 2003). Indeed, the relationship between DNA methylation and H3K9me3 seems to be highly context dependent, with loss of DNA methylation leading to reduced H3K9me3 at some loci and in some cell types but not others (Espada *et al.* 2004, Tsumura *et al.* 2006, Nguyen *et al.* 2002), and loss of H3K9me3 having similarly variable effects on the levels of DNA methylation (Lehnertz *et al.* 2003, Leung *et al.* 2014, Nakamura *et al.* 2012, Dong *et al.* 2008).

## 1.2 Chromatin Organisation and Higher Order Structure

### 1.2.1 Chromatin in the Nucleus

#### 1.2.1.1 Nuclear chromatin organisation

While we have so far discussed chromatin in terms of the linear nucleosomal structure, we must also consider the context within which chromatin is actually organised within the nucleus *in vivo*. Observations of a certain level of organisation

within the nucleus have been made since the early days of intra-cellular microscopy and include the aforementioned distinction between euchromatin and heterochromatin based on apparent domains within the nucleus (Passarge 1979). Such domain compartmentalisation is particularly evident in mouse cells, where bright foci of DNA correspond to large domains of heterochromatic H3K9me (Maison *et al.* 2010). More recent microscopy observations have revealed further levels of organisation, suggesting a highly structured nucleus. Fluorescence in situ hybridisation (FISH) experiments have shown that individual chromosomes organise into relatively distinct chromosome territories, with active genes coating the outside of the territory or extending into its "corona" to facilitate access of the transcriptional machinery (Bickmore *et al.* 2013, Cremer and Cremer 2010, Boyle *et al.* 2011). A link has also been established between the position of a region of DNA within the nucleus and its expression level. Broadly speaking, a position towards the centre of the nucleus correlates with high gene density and an active genetic profile while a position on the periphery is most associated with low gene density or silenced genes, and all the chromatin modifications that such profiles entail (Cremer and Cremer 2010). This is linked to the concept of so-called lamin-associated domains (LADs): regions with peripheral nuclear localisation achieved through interaction with proteins of the nuclear lamina (Guelen *et al.* 2008). It remains a topic of some debate whether silencing or activation of a chromatin region leads to its repositioning in the nucleus by some mechanism, or whether association with the nuclear lamina is itself the cause of silencing of the region, or if it is some combination thereof (Finlan *et al.* 2008, Demmerle *et al.* 2013, Zullo *et al.* 2012, Therizols *et al.* 2014).

More recently chromosome conformation capture (3C) technology and its derivatives 4C, 5C and Hi-C have provided a means to investigate the nuclear chromatin organisation in a distinct way, by examining the frequency of interaction between regions of DNA, and thus the likelihood that they are in close proximity in the cell (Dekker *et al.* 2002, de Laat and Dekker 2012). The majority of interactions measured by these techniques correspond to intra-chromosomal interactions, supporting the concept of chromosome territories (Lieberman-Aden *et al.* 2009). However, the smaller but significant number of inter-chromosomal interactions have also proven very informative, suggesting a certain amount of clustering of genes with similar expression profiles (Lieberman-Aden *et al.* 2009, Sexton *et al.* 2012).

These results thus also support a certain amount of compartmentalisation based on gene expression levels. PRC proteins have recently been shown to contribute to the 3D organisation of their target genes in Drosophila, thus facilitating their silencing (Entrevan *et al.* 2016). It is possible that similar processes of transcriptional silencing or activation could likewise lead to clustering of their target genes, providing some explanation for these trans-interactions.

Deep Hi-C datasets have revealed a finer level of organisation within chromosomes corresponding to regions of strong but isolated interaction which have been termed topological associated domains (TADs) (Dixon *et al.* 2012). These regions average around 100kb in *Drosophila* and 1Mb in mammals. Interaction of two regions within a TAD has a high probability, while interaction between regions from two different TADs, even if the distance between them is the same, is much less likely. TADs show a high degree of alignment with the distribution of active or repressive chromatin marks in the genome, highly suggestive of a link between nuclear organisation and transcriptional activity (Bickmore and van Steensel 2013). However, the stability of TAD boundaries between cell types and in cases of major perturbation of chromatin modifications suggests that TAD structure across a chromosomal region has some basis in the underlying gene sequence (Dixon *et al.* 2012, Nora *et al.* 2012).

Primary candidates for the organising factors behind TAD and broader nuclear organisation are the insulator protein CTCF and the cohesin complex. CTCF has long been known to have a role in compartmentalising the genome, for instance by blocking communication of enhancer elements outside of their intended domain (Phillips and Corces 2009). Binding sites for CTCF are enriched in the boundary regions of TADs, although these boundary sites only account for a small proportion of the total CTCF-binding sites (Dixon *et al.* 2012). Cohesin is one of the similar SMC complexes which have a ring-like structure (**Figure 1.3**) (Nasmyth and Haering 2009). In cohesin this ring is composed of the two long SMC proteins SMC1 and SMC3, which dimerise via their hinge domains to form a “V” shape. The ring is then “capped” by the kleisin family protein RAD21 (budding yeast SCC1), supported by the HEAT repeat-containing proteins STAG1 and STAG2 (also called SCC3 in budding yeast, psc3 and rec11 in *S. pombe*, SA and SA-2 in *Drosophila*). The best-known function of the cohesin complex is in the cohesion of sister chromatids during

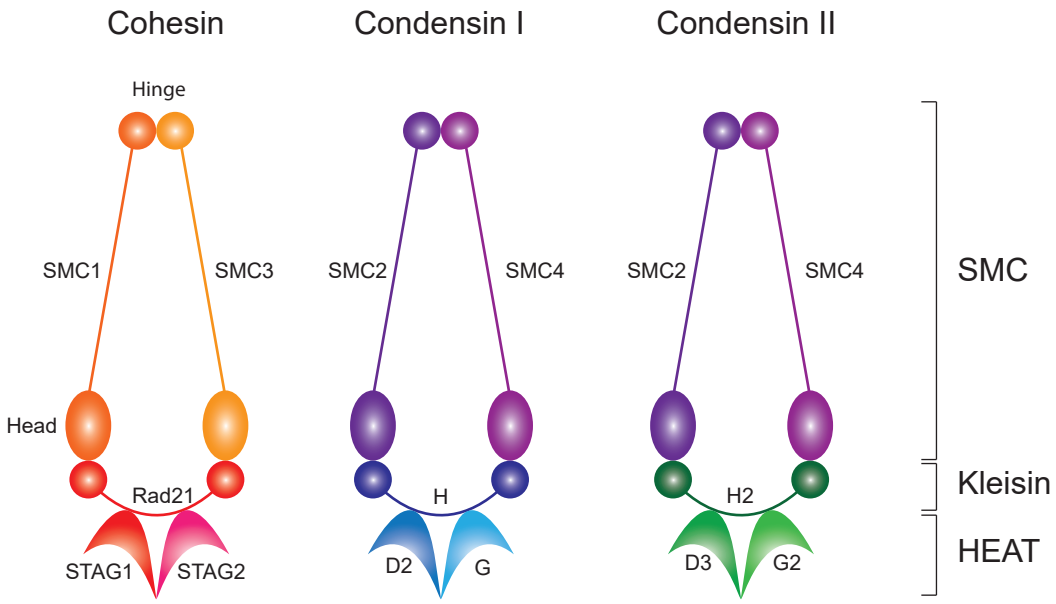


**Figure 1.3 - Structure and composition of the SMC protein complexes**

Structure of the chromatin organising SMC protein complexes cohesin, condensin I and condensin II. Each is composed of two SMC family proteins that dimerise in a “V” shape, capped by a kleisin protein and two HEAT repeat-containing subunits.

mitosis, which will be discussed later. More recently however a role for cohesin in organising interphase chromatin has emerged. CTCF and cohesin were found to colocalise extensively in the mammalian genome and to have some interdependence of localisation and function (Wendt *et al.* 2008, Parelho *et al.* 2008, Rubio *et al.* 2008). Other results found that both contribute differentially to the organisation and stability of TADs, although other processes are also believed to be at work (Sofueva *et al.* 2013, Zuin *et al.* 2013). Based on this, it has been proposed that CTCF and cohesin work together to organise nuclear chromatin (Philips and Corces 2009).

The primary model for TAD formation by CTCF and cohesin is the loop-extrusion model of chromatin organisation (Alipour and Marko 2012, Sanborn *et al.* 2015, Busslinger *et al.* 2017). In this model cohesin binds to chromatin and uses its ATPase activity to push a loop of chromatin through its ring. Different versions of the model have been proposed whereby either a whole loop is pushed through a single cohesin complex, or only one DNA strand is pushed through the cohesin ring and looping is achieved through cohesin dimerisation (**Figure 1.4**). Chromatin will continue to be pushed through the ring, making a larger and larger loop, until either that cohesin ring dissociates from the chromatin (possibly through the action of the Wapl protein) or until the cohesin ring reaches a boundary, defined by the presence of CTCF, or another cohesin complex, neither of which it can pass. In this manner, chromatin is organised into domains of similar features simply through the definition of loop boundaries and the action of cohesin as an organiser. Presence of loops within loops can assure the stability of these domains, with several cohesin rings bound to the base of the same loop ensuring that the dissociation of one does not lead to the disappearance of that loop.

#### 1.2.1.2 Nuclear Chromatin Higher Order Structure

Another important question concerning the *in vivo* organisation of chromatin is whether it adopts a recognisable structure beyond the level of the 10nm beads-on-a-string fibre. As discussed above, a number of histone modifications are believed to result in structural changes to the chromatin fibre, leading to compaction of silent heterochromatin and opening up of active euchromatin, and it has been proposed

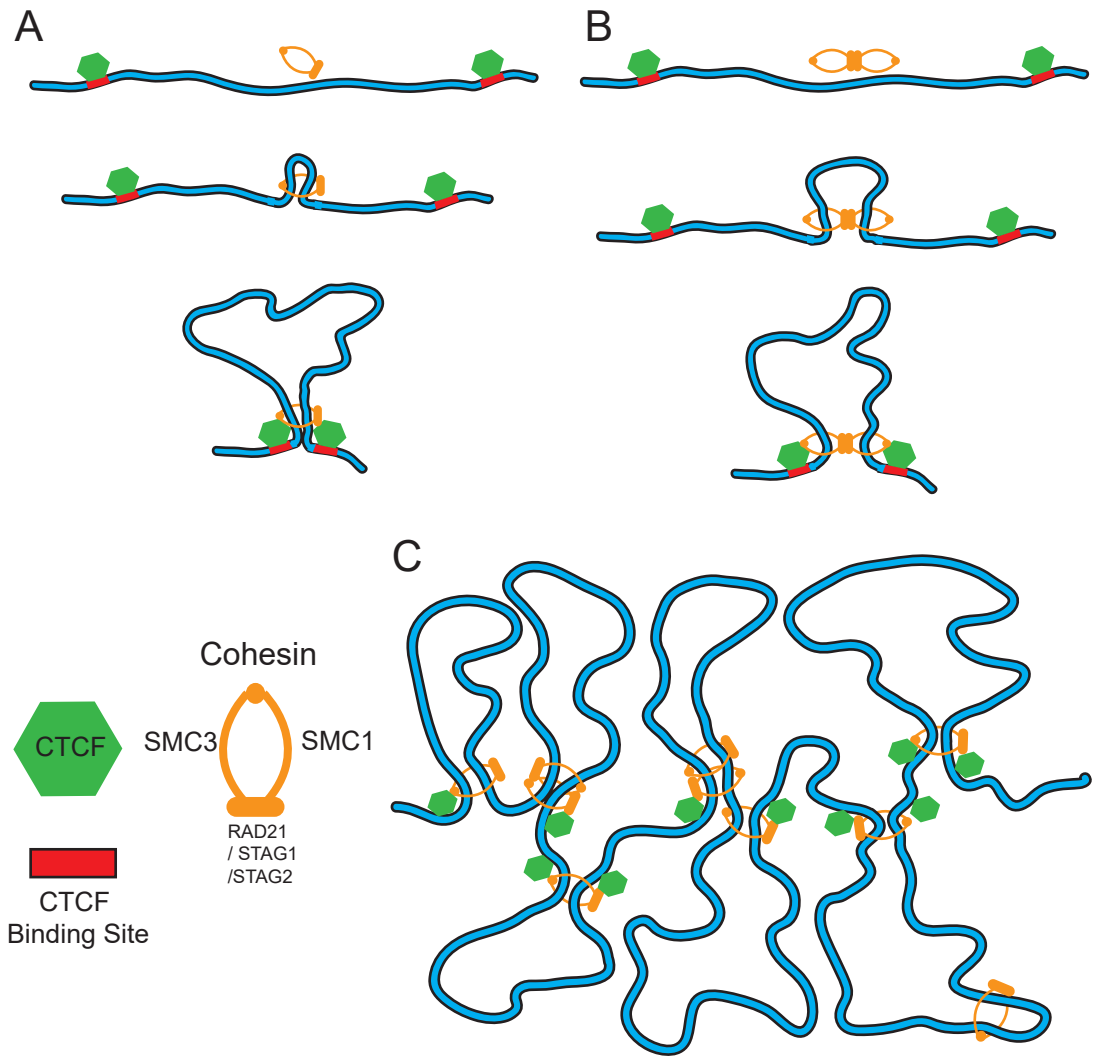


**Figure 1.4 - The loop extrusion model of chromatin organisation**

Graphical representation of potential models of chromatin organisation by loop extrusion in the nucleus. In these models either one (**A**) or a pair (**B**) of cohesin complexes bind to DNA and push it through their rings, thus extruding a loop. The loop will continue to grow until the cohesin reaches an obstacle, such as another cohesin or a CTCF protein, or falls off (not shown). In this way chromatin can be organised into successive loops (**C**) potentially explaining the existence of TADs (contact within loops) and chromosome contact between regions of DNA that are distant in the linear genome (between loops). Generation of loops within loops could provide robustness in the event of one cohesin detaching from the DNA.

that such compaction can lead chromatin being reorganised into higher order 30nm fibres.

Evidence for a 30nm fibre of chromatin first came from transmission electron microscopy (EM) experiments of purified nucleosomes *in vitro* (Finch and Klug 1976). They proposed a solenoidal or “one-start helix” model for chromatin fibres. Later a zig-zag or “two-start helix” model was proposed, with further evidence coming from the X-ray structure of a tetranucleosome (Woodcock *et al.* 1984, Schalch *et al.* 2005). These two models of folding have dominated the field, however variations and other models have been proposed (van Holde and Zlatanova 2007) (**Figure 1.5**).

Whether the 30nm fibre exists *in vivo* is the subject of much debate. It has been pointed out that the primary models for the 30nm fibre require a high level of uniformity in DNA linker length, which is simply not observed in cells (Routh *et al.* 2008, van Holde and Zlatanova 2007). The best evidence for a 30nm fibre *in vivo* comes from analysis of echinoderm sperm cell and chicken erythrocyte chromatin, highly transcriptionally inactive cells that are unlikely to be representative of the situation in more dynamic nuclei (Horowitz *et al.* 1994, Woodcock 1994). Conversely, new imaging techniques have found no evidence for a 30nm fibre in other nuclei, even in regions of heterochromatin (Ahmed *et al.* 2010, Fussner *et al.* 2011, Ricci *et al.* 2015) and Hi-C data has also been shown to be more consistent with nuclear chromatin being a highly flexible assemblage of 10nm fibres (Sanborn *et al.* 2015).

Fussner *et al.* (2011) make an important distinction between chromatin “compaction” and “condensation”. They point out that chromatin organised into a 30nm fibre would take up a larger volume than chromatin in a 10nm fibre, just as the 10nm fibre has a much greater volume than naked DNA. They argue that the point of chromatin, aside from balancing the negative charge of DNA, is to allow it to become highly folded, and that in this respect the 10nm fibre is as capable of achieving high levels of compaction as the 30nm fibre. In light of this proposed superfluousness of the 30nm fibre and the increasing uncertainty over its existence, or at least widespread presence, *in vivo*, many models now seek to explain chromatin structure and organisation without recourse to it. Such is the case for the loop-extrusion model of

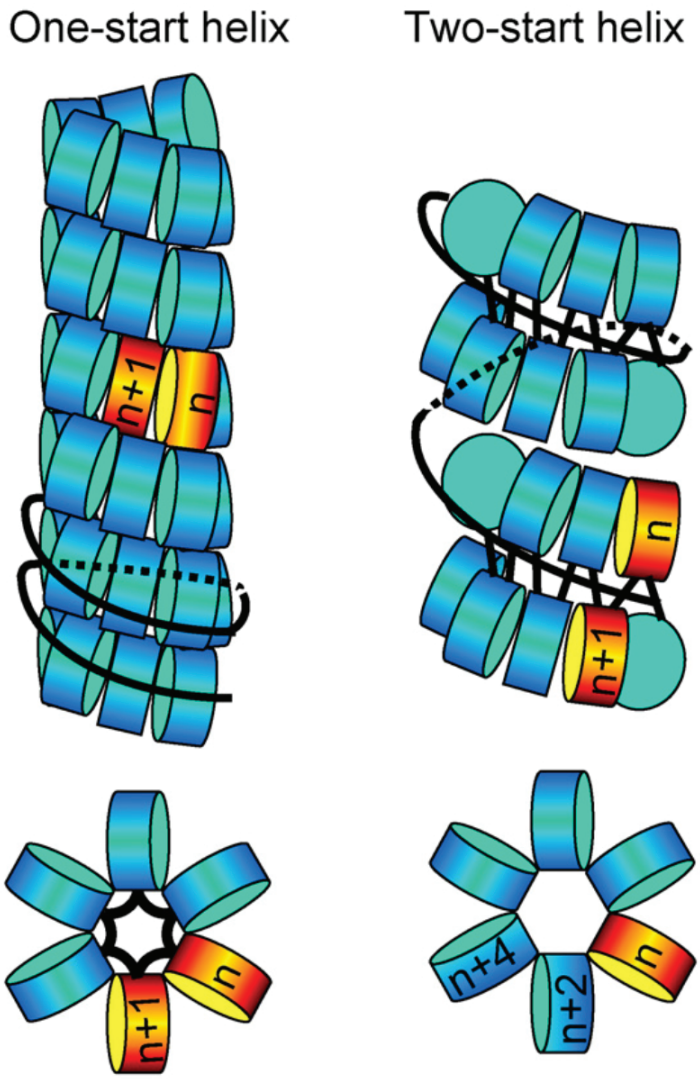


**Figure 1.5 - Primary models of 30nm chromatin fibre structure**

Side and top views of the two main models of chromatin organisation into 30nm fibres: a one-start helix (solenoid) and two-start helix (zig-zag). Nucleosomes are represented by blue cylinders, with two linearly sequential nucleosomes highlighted in orange in each case to show their different spatial localisation and nucleosome neighbours.
From van Holde and Zlatanova 2007.

TAD formation discussed above and is also true for many recent models of mitotic chromosome structure, which will be discussed next.

### 1.2.2 The Mitotic Chromosome

#### 1.2.2.1 Definition and Function

In all eukaryotes, chromatin becomes highly condensed during mitosis, although the degree of condensation varies across species (Belmont 2006). In mammalian cells, this corresponds to a decrease in volume of about 2-3 fold compared to interphase chromatin (Vagnarelli 2012). The reasons for this condensation become clear when the process of cell division is examined (**Figure 1.6**).

Chromosomes begin to condense in prophase of mitosis, just prior to nuclear envelope break down (NEB) in most organisms. By late prophase they have adopted a highly reproducible cylindrical structure, with the two sister chromatids held together by cohesin at the centromere. The centromere is a ubiquitous chromosomal feature of eukaryotes and in metazoa and some unicellular organisms is generally composed of large tandem repeat arrays of DNA with constitutively heterochromatic profiles. Its primary function is to serve as a platform for kinetochore assembly, the machinery which coordinates the movement of the chromosomes in mitosis through microtubule attachments. The individual chromosomes are aligned on the metaphase plate through the antagonistic forces of the microtubules emanating from the opposite poles of the dividing cell. Once the cell cycle checkpoints are satisfied, cohesin is cleaved at anaphase onset and the sister chromatids are pulled towards opposite poles, at which point the cell itself divides along the former metaphase plate in telophase, the nuclear envelope reforms, and the chromosomes begin to decondense (Pollard and Earnshaw).

The role of mitotic chromosome formation is thus primarily two-fold. First it is essential that chromosomes maintain individual identity during mitosis so that they can be segregated accurately, with one and only one copy of each homologue ending up in each daughter cell. Chromatin in its jumbled interphase state must thus be decatenated so that any potential loops or contacts between different chromosome fibres are removed before mitosis, lest the pulling forces in anaphase

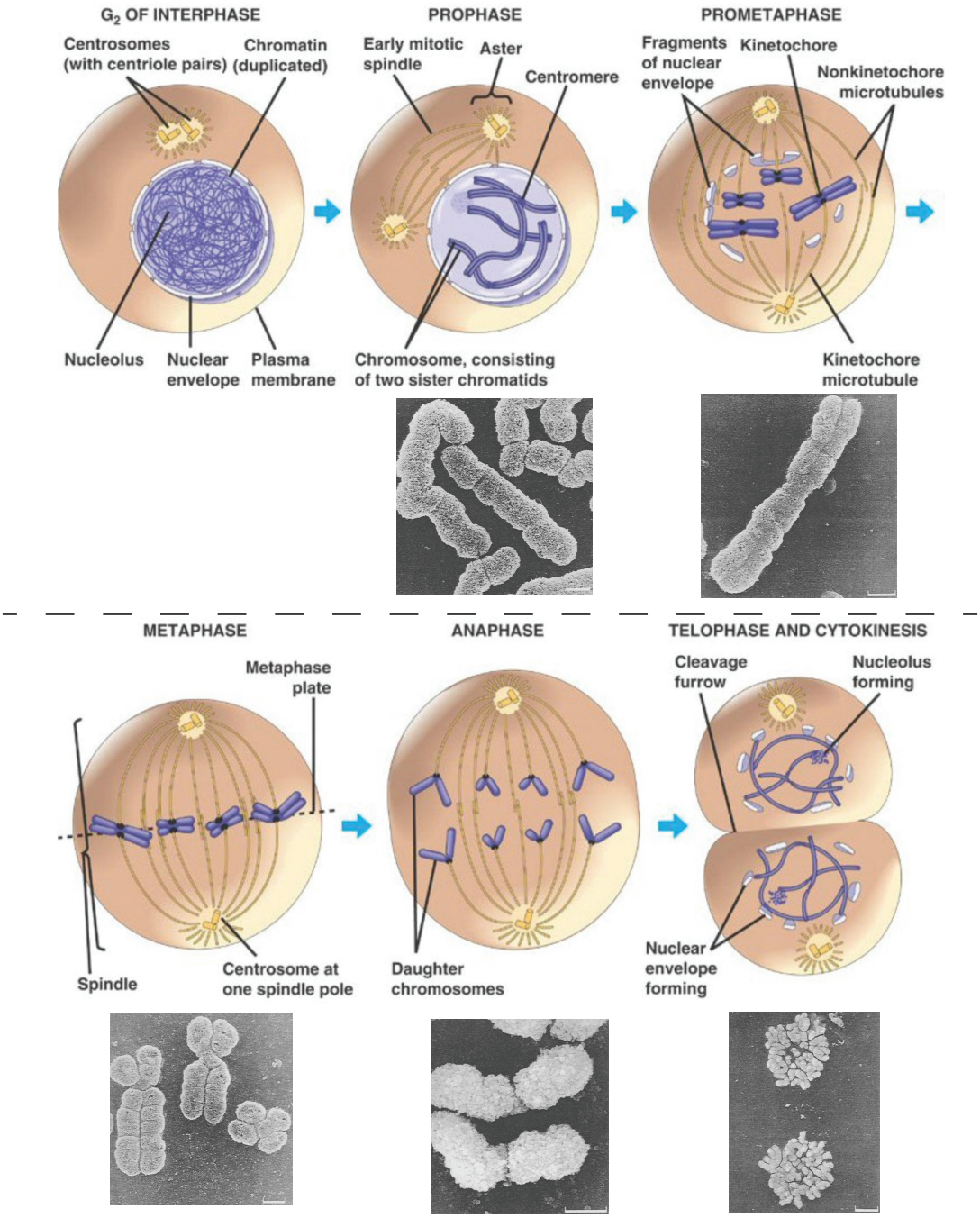


**Figure 1.6 - Chromosome condensation in the stages of mitosis**

Schematic representation of cell cycle stages with electron micrographs of chromosome morphology below (from Sumner 1991). In interphase chromatin is in its decondensed state. Chromosome condensation begins in early prophase such that by the time of nuclear envelope breakdown chromosomes are individualised in a long, thin form. Condensation proceeds in prometaphase and metaphase to form compact cylindrical structures and continues into anaphase when the chromosomes reach their highest state of compaction. Once the cell has divided the chromosomes begin to decondense as the nuclear envelope reforms around them. Scale bars represent 5μm for telophase, 1μm otherwise.

lead to shearing or missegregation. Secondly, the individual chromosomes, which in most eukaryotes are only being pulled and organised from a single point along their length, must remain entirely within the volume of the daughter cell (and its newly forming nucleus) towards which its centromere is being pulled. In the absence of this condensation chromosomes would extend past the metaphase plate and be cleaved by the cytokinetic apparatus. The importance of this role is further suggested by the observation that chromosomes reach their highest level of compaction in anaphase (Mora-Bermudez *et al.* 2007) and by the increase in segregation defects in cells defective in condensation (Nasmyth 2002).

The necessity of forming individualised and condensed mitotic chromosomes is thus evident. What is remarkable about mitotic chromosomes, particularly in humans and other mammals, is the reproducibility of their structure and features. For instance, various staining techniques, most commonly giemsa staining, reveal reproducible banding patterns of differential staining along chromosomes that enable identification of individual chromosomes (Caspersson *et al.* 1971, Sumner 1982). Within a single species the width and length of individual chromosomes also remain remarkably reproducible. And, regardless of their position, centromere regions appear as distinct regions on mitotic chromosomes, as constricted chromatin regions with decreased width (Blackburn and Szostak 1984). This is particularly obvious in extreme cases such as the large centromere regions of muntjac chromosomes, which show a reduced diameter along their entire extended length (Korf and Diacumakos 1978, Hsieh *et al.* 2014). The reasons for this distinct structure remain unclear, but any model of mitotic chromosome structure must also explain such major structural differences along their length. Below I highlight some factors that are, or have been thought to be, involved in mitotic chromosome folding and structure before discussing various models that seek to explain these processes with reference to recent structural data.

#### 1.2.2.2 Factors involved in mitotic chromosome structure

##### 1.2.2.2.1 30nm chromatin fibres in mitosis

As mentioned above, evidence has been accumulating that the organisation of chromatin into a 30nm fibre is not widespread in nuclei and unnecessary to explain

the levels of organisation observed. However, the evidence discussed thus far did not address the possibility that 30nm fibres are present within mitotic chromosomes. Similar to the situation in interphase cells, the discovery of the 30nm fibre *in vitro* led to a long-held assumption that, as a more compact structure, the 30nm fibre would form the basis of even higher levels of chromosomal organisation, ultimately leading to the final mitotic chromosome structure (Kireeva *et al.* 2004). However, the lack of evidence for a 30nm fibre *in vivo* has shifted thinking and models in new directions (Maeshima *et al.* 2014). Early evidence for the lack of a 30nm fibre came from cryo-electron microscopy (Cryo-EM) of native mammalian mitotic chromosome sections. In two separate studies, the second correcting for a potential source of signal masking, chromosomes were found to be composed of regular spaced chromosome fibres of around 11nm, with no detection of larger 30nm structures (Dubochet *et al.* 1986, Eltsov *et al.* 2008). Long-standing evidence for the presence of 30nm fibres came from small-angle X-ray scattering (SAXS) of mitotic chromosomes (Langmore and Paulson 1983), which showed regularly spaced features of chromatin at 6, 11 and 30nm, which the authors argued resulted from the 3D packing of nucleosomes in a face-to-face, edge-to-edge and side-by-side manner, respectively, consistent with a densely packed 30nm fibre. However, these results more recently shown to be affected by contamination with ribosomes on the surface of chromosomes, regularly spaced at intervals of around 30nm (Nishino *et al.* 2012). Removal of ribosomes from similar preparations resulted in the disappearance of the 30nm signal, while the 6 and 11nm signals resulting from the internal structure of nucleosomes remained unchanged. As in interphase then, there is little evidence for 30nm chromatin fibres contributing to mitotic chromosome structure *in vivo*, while much of the previous evidence has been overturned. While some models do take into account the possibility of a 30nm structure, most prefer not to rely on it to explain the compaction of mitotic chromosomes, simply using the well-defined 10nm fibre as the basic element from which mitotic chromosomes are formed.

#### 1.2.2.2.2 Proteins regulating chromosome architecture

Current estimates suggest that non-histone chromatin proteins account for around 40% of mitotic chromosome protein levels (Ohta *et al.* 2010). Not much is known about these proteins, however key roles for a few specific non-histone proteins and protein complexes have been elucidated. The best studied of these are the

condensin I and II complexes. These complexes are related to the previously mentioned cohesin complex and include the structural maintenance of chromosomes proteins SMC2 and SMC4 along with the cap proteins CAPH1, CAPG1 and CAPD2 (for condensin I) or CAPH2, CAPG2 and CAPD3 (for condensin II) (Hirano 2012) (**Figure 1.3**). They were initially identified from purification of the so-called chromosome "scaffold", a structure that is left behind upon histone depletion of mitotic chromosomes (Paulson and Laemmli 1977, Lewis and Laemmli 1982). It is believed that the presence of both condensin complexes is an ancestral trait of eukaryotes, though condensin II has been lost in several lineages including fungi (Hirano 2012).

The central importance of the condensin complexes is apparent from the defects caused by their removal or depletion, although the effects of these experiments vary from species to species and the relative role of the two complexes is not fully resolved. In both *S. cerevisiae* and *S. pombe* deletion of one of the subunits of the single remaining condensin complex leads to chromosome condensation and segregation defects (Strunnikov *et al.* 1995, Saka *et al.* 1994). In most vertebrates, reduction of either condensin complex results in severe cellular defects, and an even more severe phenotype is observed when both are depleted (Wignall *et al.* 2003, Ono *et al.* 2003, Hirota *et al.* 2004). Chicken DT40 chromosomes on the other hand show only modest phenotypes for condensin depletion under physiological salt conditions, although condensation was delayed and hypotonic conditions revealed them to be structurally abnormal (Hudson *et al.* 2003). Thus, though condensins are undoubtedly involved and highly important in mitotic chromosome compaction, certain experiments highlight the existence of other, independent pathways involved, leading some to propose an additional unknown "regulator of chromosome architecture" (RCA) factor regulated by the targeting of protein phosphatase 1 by the Repo-man protein (Vagnarelli *et al.* 2006).

The two condensin complexes have been found to localise axially down the centre of mitotic chromosomes in spatially distinct domains (Hirota *et al.* 2004, Ono *et al.* 2003). Interestingly, changes in the ratio of condensin I to condensin II have been found to lead to changes in mitotic chromosome shape. This has been achieved both through assembly of mitotic chromosomes *in vitro* in Xenopus cell extracts when the ratio of condensins supplied was altered (Shintomi and Hirano 2011), and

by *in vivo* inactivation or hyperactivation of one of the two complexes (Lai *et al.* 2010, Yamashita *et al.* 2011). In all cases, an excess of condensin II led to a shorter, fatter chromosome structure, while an excess of condensin I led to a longer, thinner structure, suggesting different roles in the 3D compaction of chromatin during mitotic chromosome formation.

These findings have been coupled with cytological observations for a chronological model of their respective roles in the cell cycle. In interphase condensin I is cytoplasmic while condensin II is nuclear (possibly entailing a further role in interphase nuclear chromatin organisation) (Steen *et al.* 2000, Schmiesing *et al.* 2000, Hirota *et al.* 2004). Given that chromatin begins to compact in early prophase, prior to NEB, it follows that condensin I cannot be involved in the early stages of chromatin compaction, but will only gain access to chromatin in late prophase after NEB. It has thus been proposed that condensin II mediates the initial compaction of chromatin into cylindrical structures and that condensin I only later gains access to the already individualised chromosomes, which it proceeds to further compact diametrically (Antonin and Neumann 2011, Kschonsak and Haering 2015) (**Figure 1.7**).

Other proteins involved in mitotic chromatin structure include Topoisomerase IIα (TopoIIα) and the microtubule motor protein Kif4, aka chromokinesin. TopoIIα has also been shown to be enriched in chromosome scaffold preparations (Earnshaw *et al*. 1985). TopoIIα and condensin localise along the chromosome axis in an alternating pattern (Maeshima and Laemmli 2003) and depletion of TopoIIα also leads to chromosome segregation defects (DiNardo *et al.* 1984, Uemura *et al.* 1987, Holm *et al.* 1985), but its role in mitotic chromosome formation remains unclear. Being a member of the topoisomerase superfamily of proteins frequently involved in the decatenation or unwinding of DNA (Vos *et al.* 2011), a role in resolution of the individual chromosomes in mitosis has long been associated with TopoIIα (Coelho *et al.* 2003). However, many studies also ascribe to TopoIIα a role in the structural organisation or integrity of mitotic chromosomes (Chang *et al.* 2003, Samejima *et al.* 2012). A key aspect of this role may lie in relieving the positive supercoiling of DNA introduced by condensin-mediated organisation (Baxter *et al.* 2011). Kif4 also localises to the chromosome axis in mitosis, along with other locations (Aizawa *et al.* 1992, Vernos *et al*. 1995; Lee *et al.* 2001). Kif4 appears to have many functions in

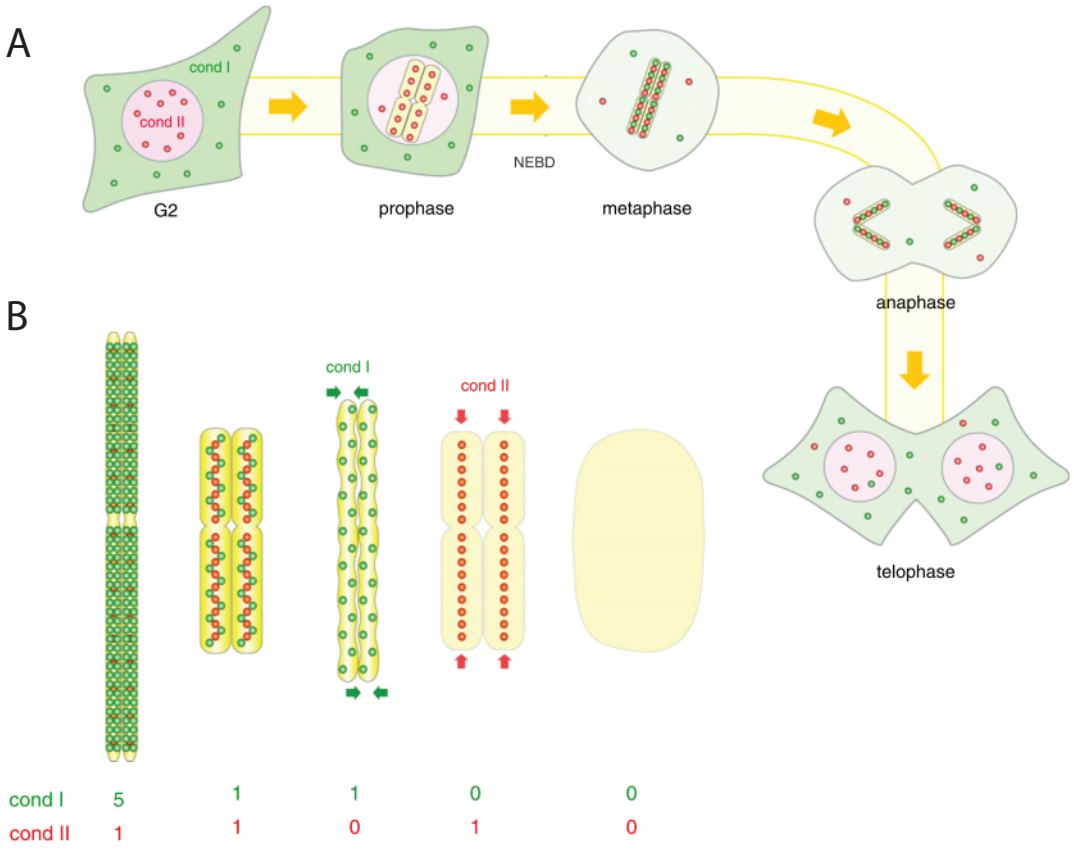


**Figure 1.7 - Relative roles of condensin I and condensin II in mitosis**

**A** - Localisation of condensin I (green) and II (red) at different stages of the cell cycle. Early chromosome compaction is mediated by condensin II only while condensin I gains access to the chromosomes after nuclear envelope breakdown in late prophase.
**B** - Morphology of mitotic chromosomes in several natural or artificially induced situations in which the ratio of the two condensin complexes varies, revealing the different role each has in shaping the chromosome.
Adapted from Hirano 2012.

mitosis, however its depletion from mitotic cells resulted in a similar defect to the removal of condensin I and has thus been proposed to work in parallel or in concert with condensin-mediated compaction (Samejima *et al.* 2012). Some models suggest that TopoIIα and Kif4 may be involved more in the chromatin compaction of the later stages of the cell cycle, culminating in anaphase when the chromosome is at its most compact, thus working independently but in an additive manner to condensin-mediated compaction (Samejima *et al.* 2012, Kchonsak and Haering 2015).

The cohesin complex may also have a role in chromosome structure beyond its well-known function of sister chromatid cohesion. As previously described, cohesin is essential in ensuring that sister chromatids remain paired until the onset of anaphase (Nasmyth and Haering 2009). Cohesin is localised quite broadly along the length of chromosomes in early mitosis but most is gradually lost from chromosome arms during prophase and prometaphase, remaining primarily at centromeres where it mediates sister centromere cohesion until cleaved by separase at anaphase onset (Waizenegger et al. 2000, Hauf et al. 2001). The effect cohesin has on the mitotic chromosome structure is usually explained as an indirect consequence of its influence on the localisation or activity of condensin (Hirota *et al.* 2004, Shintomi and Hirano 2011). Notably, removal of cohesin from chromosome arms was found to coincide with loading of extra condensin II but not condensin I, thus altering the ratio of the two condensins (Shintomi and Hirano 2011). Such a change over the arms but not at centromeres, where cohesin is maintained, might partly explain the observed difference in structure between centromere regions and chromosome arms on metaphase chromosomes.

A holistic view of mitotic chromosome folding would have to include all of these factors, and likely others, working in concert. Examination of the individual relationships between these proteins have revealed various instances of interdependence of localisation as well as both cooperative and antagonistic activities, with some level of redundancy, all of which likely contributes to the robustness and reproducibility of mitotic chromosome structure during every mitosis.

#### 1.2.2.2.3 DNA Sequence determinants of chromosome architecture

Early evidence suggested that DNA sequence might play an important role in organising chromatin along mitotic chromosomes. Certain DNA sequences, termed Scaffold Attachment Regions (SARs), were identified as being enriched in chromosome scaffold preparations (Merkovitch *et al.* 1988). These SARs were found to be enriched in consensus recognition sequences for Topoisomerase II proteins (Gasser and Laemmli 1986, Gasser *et al.* 1986), and TopoIIα was shown to bind SARs *in vitro* (Adachi *et al.* 1989). These observations led to the suggestion that structural proteins, such as condensin and TopoIIα, might organise chromatin into loops anchored at such sequence-defined recognition sites, thus providing a mechanistic basis for regular organisation of chromatin in mitosis. In contrast, the recent implementation of Hi-C technology has found no evidence for specific sequences playing any role in the anchoring of chromatin loops (Naumova *et al.* 2009, Gibcus *et al.* 2018). Rather the interaction patterns of DNA within a population of mitotic chromosomes were highly homogeneous, suggesting chromatin is organised in a stochastic manner regardless of underlying sequence. Chromatin immunoprecipitation (ChIP) of condensin also failed to identify a sequence bias for condensin positioning along chromatin (Gibcus *et al.* 2018).

#### 1.2.2.2.4 Histone modifications in mitosis

The observation that a certain level of chromatin compaction proceeds in the absence of the known organising proteins has led some to make a distinction between the processes of chromosome “compaction” and chromosome “shaping” (Zhiteneva *et al.* 2017). In the absence of other candidates, it has been suggested that the proposed RCA pathway of chromatin compaction might be mediated directly by chromatin interactions through the effect of histone modifications (Kschonsak and Haering 2015). Indeed, chromatin isolated from mitotic cells has been shown to aggregate more readily than interphase chromatin, even in the absence of “shaping” proteins such as condensin, cohesin, TopoIIα and Kif4 (Zhiteneva *et al.* 2017).

The best studied mitotic histone modification is the highly conserved phosphorylation of histone H3 at serine 10. H3S10 is widely phosphorylated in a cell cycle dependent manner by Aurora B, a well-known cell cycle kinase, beginning at

centromeres in G2 phase and spreading out to cover the whole chromosome in prophase (Prigent and Dimitrov 2003). Through steric effects this phosphorylation leads to dissociation of HP1 from the adjacent methylated H3K9, which is required for proper condensation of chromatin (Fischle *et al.* 2005, Hirota *et al.* 2005). It has been suggested that targeting of Aurora B is achieved in part through a DNA methylation-mediated mechanism, which would indeed help promote phosphorylation at heterochromatic regions where HP1 binds (Monier *et al.* 2007).

Both DNA and histone modifications can thus have important roles to play in chromatin condensation by similar mechanisms to their functionality in interphase, through promotion or inhibition of recruitment of secondary factors. Some evidence also suggests that histone modifications might have a more direct effect on mitotic chromosome structure, potentially by the same electrostatic mechanisms that affects interphase chromatin. H4K16ac has been shown to abrogate higher order nucleosome interactions *in vitro* (Shogren-Knaak *et al.* 2006), and its deacetylation in mitosis has been suggested as another function of H3S10ph through recruitment of the deacetylase Hst2p (Wilkins *et al.* 2014). Histone acetylation generally has been found to be reduced in mitotic chromatin compared with interphase, and other chromatin marks, including H3T3ph, H3S28ph, H2AT120ph, H3K9me3 and H2BK120 de-ubiquitination have been shown to be enriched in mitotic chromatin and have thus been proposed as potential actors in chromosome condensation, though no mechanism for their action has yet been elucidated (Zhiteneva *et al.* 2017, Goto *et al.* 2002**,** McManus *et al.* 2006, Park *et al.* 2011).

One group of modifications which may combine these mechanisms to have a strong impact on mitotic chromatin structure is the phosphorylation of the linker histone H1. H1 is phosphorylated at a number of residues in a cell cycle dependent manner. Overall phosphorylation of H1 is lowest in G1 phase and increases gradually to its maximal level in metaphase before dropping off sharply thereafter (Talasz *et al.* 1996). CDKs, particularly CDK1, have been implicated in the phosphorylation of many of these residues, but other kinases including Aurora B and protein kinase A are also involved (Izzo and Schneider 2016). Inhibition of H1 phosphorylation in mouse and Chinese hamster ovary cells prevented proper mitotic chromosome condensation and entry into mitosis, revealing the crucial role of this modification in mitotic chromatin organisation (Th'ng *et al.* 1994, Gurley *et al.* 1975, Izzo and

Schneider 2016). Possible mechanisms of action of the phosphorylation include changes in the secondary structure of H1 CTDs, inhibition of chromatin remodellers or, similar to H3S10ph, preventing the binding of HP1 to chromatin (Roque *et al.* 2008, Horn *et al.* 2002, Daujat *et al.* 2005).

#### 1.2.2.3 Models of mitotic chromosome structure

##### 1.2.2.3.1 Hierarchical folding models

The original observation of the *in vitro* folding of the 10nm fibre into a regular 30nm fibre suggested the possibility of even higher orders of chromatin organisation. This led to the proposal that mitotic chromosomes were the final, highest level of organisation achieved through stepwise folding of chromatin into successively larger fibres in a hierarchical manner (Zatsepina *et al*., 1983; Belmont *et al*., 1987, Woodcock and Gosh 2010). However, this model has become increasingly incompatible with recent observations, including the mechanical properties of chromosomes, and has failed to provide an explanation for the axial localisation of scaffold proteins (Poirier and Marko 2002, Maeshima and Laemmli 2003). In an attempt to reconcile all of these observations, a model has been proposed in which chromosomes first fold in a hierarchical manner into fibres of 30, 100-130, 200-250 and 500-750nm in diameter (Kireeva *et al.* 2004). After this initial folding, condensins and other scaffold proteins are then proposed to act as a "glue" that holds the structure together, localising to the centre of the chromosome and binding different chromatin fibres to maintain structural integrity (**Figure 1.8 A**). However, this model still does not account for the observation that condensin is required relatively early in the process of chromatin condensation, not simply as a passive structural element (Kschonsak and Haering 2015). With the increasing scepticism concerning the existence of a 30nm fibre *in vivo* and the Hi-C data that does not fit with a hierarchical organisation of chromatin (Naumova *et al.* 2009, Gibcus *et al.* 2018), the prevailing view of mitotic chromosome structure is beginning to abandon such models.

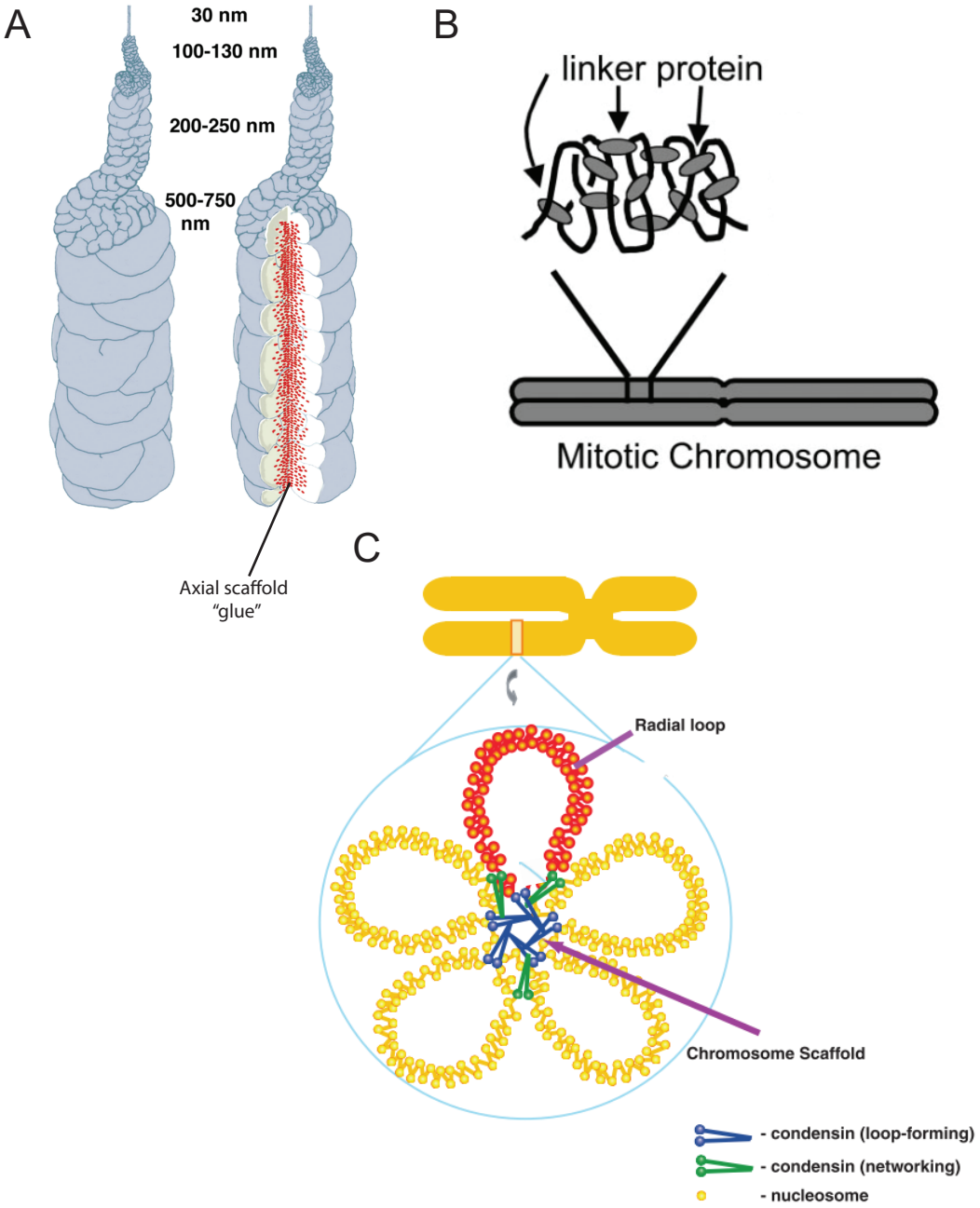


**Figure 1.8 - Major models of mitotic chromosome structure**

Simplified representations of the primary models for mitotic chromosome structure:
**A** - The hierarchical folding model, in which chromatin folds in fibres of increasing diameter, with the inclusion of an axial protein "glue" to explain the role of the proposed scaffold proteins.
**B** - The network model, in which chromatin fibres are more disorganised and kept together through crosslinking by linker proteins
**C** - The radial loop model, in which loops of chromatin are organised by attachment at their base to an axial protein scaffold from which they radiate out.
Adapted from Kireeva et al. 2004, Poirier and Marko 2002 and Maeshima and Eltsov 2008.

#### 1.2.2.3.2 Network models

Strong evidence against a hierarchical organisation of chromatin in mitosis has been obtained from micromechanical analyses (Poirier and Marko 2002, Marko *et al.* 2008). Stretching an intact mitotic chromosome between two micropipettes revealed a high degree of chromatin elasticity, with the chromosome able to return to its original size even after being stretched to five times its length. The study concluded that these mechanical properties were inconsistent with a hierarchical structure, which would gradually unravel under tension. Digestion of the chromatin with restriction enzymes led to increased fragility of the chromosome correlating with the frequency of cut sites for the enzyme, further arguing against the organisation of chromatin around a rigid protein scaffold. Based on these results the authors proposed that chromatin is formed of a more disorganised network of chromatin fibres crosslinked on average every 15kb (**Figure 1.8 B**). This model is compatible with observed mechanical properties, is equally applicable in the absence or presence of 30nm fibres and was found to fit the initial mitotic Hi-C data (Naumova *et al.* 2009). However, if such a network is truly as disorganised as suggested, the highly reproducible organisation of the mitotic chromosome is difficult to explain, as is the consistent axial localisation of chromosome scaffold proteins. Furthermore, it does not allow for much variation in the roles of the key protein crosslinkers, and it is thus difficult to explain the clearly distinct roles of certain scaffold components and their effects on chromosome morphology, such as condensin I and II (Hirota *et al.* 2004, Shintomi and Hirano 2011). Nevertheless, the micromechanical data remains an important aspect of chromosome structure that must be accounted for by any model of mitotic chromosome structure.

#### 1.2.2.3.3 Radial loop models

A radial loop organisation of chromatin has been a persistent concept that, despite strong limitations that have become clear since its formulation, has been successfully updated in several respects such that it fits with a considerable amount of recent data. In its original form the idea arose from the observation and isolation of a protein scaffold that remained and retained the shape of the mitotic chromosome when most of the chromatin was stripped away (Paulson and Laemmli 1977). This initially led to the idea of a rigid protein scaffold that assembled into a

“skeleton” of the mitotic chromosome and upon which chromatin was assembled, possibly through attachments at regularly spaced, sequence determined scaffold attachment regions (Mirkovitch *et al.* 1988, Maeshima and Eltsov 2008) (**Figure 1.8 C**).

As described in **Section 1.2.2.2.3**, the concept of specific sequences that act as SARs has been increasingly discarded, and that of a rigid self-assembled scaffold no longer fits the body of evidence. Rather, to fit the Hi-C and micromechanical data a much more stochastic and dynamic model than initially envisaged is required that nonetheless explains the reproducible and stable chromosome structure observed. In order to find a mechanism for such an organisation, the loop extrusion model, already mentioned in the context of interphase chromatin organisation, was adapted to the organisation of mitotic chromatin (Goloborodko *et al.* 2016). By varying the parameters, namely residence time, processivity and the formation of nested loops, the authors were able to show that the action of simple loop extruding factors was sufficient to explain the level of compaction and observed cylindrical structure of mitotic chromosomes, even in the absence of 30nm fibres.

These initial modelling and Hi-C results for stochastic radial loops were recently extended to a more complete model by incorporating cell cycle dynamics and roles for the different scaffold proteins (Gibcus *et al.* 2018). In this study the authors propose a nested-loop structure whereby condensin II initially forms large loops of chromatin which increase in size as mitosis progresses, but are then further subdivided into smaller loops-within-loops by condensin I (**Figure 1.9**). This model not only fits with the known order of action of the two condensins but is also easily explained by loop extrusion if the residence times and/or processivity of the two condensin complexes differ. Analysis of their dynamic Hi-C data suggested a gradual helical coiling of the axial condensins, mediated by condensin II, leading to a “spiral staircase” organisation of chromatin loops. This model could explain the proposed role of condensin II in axial shortening of chromatids in later mitosis, and the longer, narrower structure of chromosomes in cells depleted of condensin (Shintomi and Hirano 2011). The mechanism of the coiling is not explained, however this does leave the model open to inclusion of other factors whose role is not yet fully clear such as Kif4, which is known to work in concert with condensin II (Samejima *et al.* 2012).

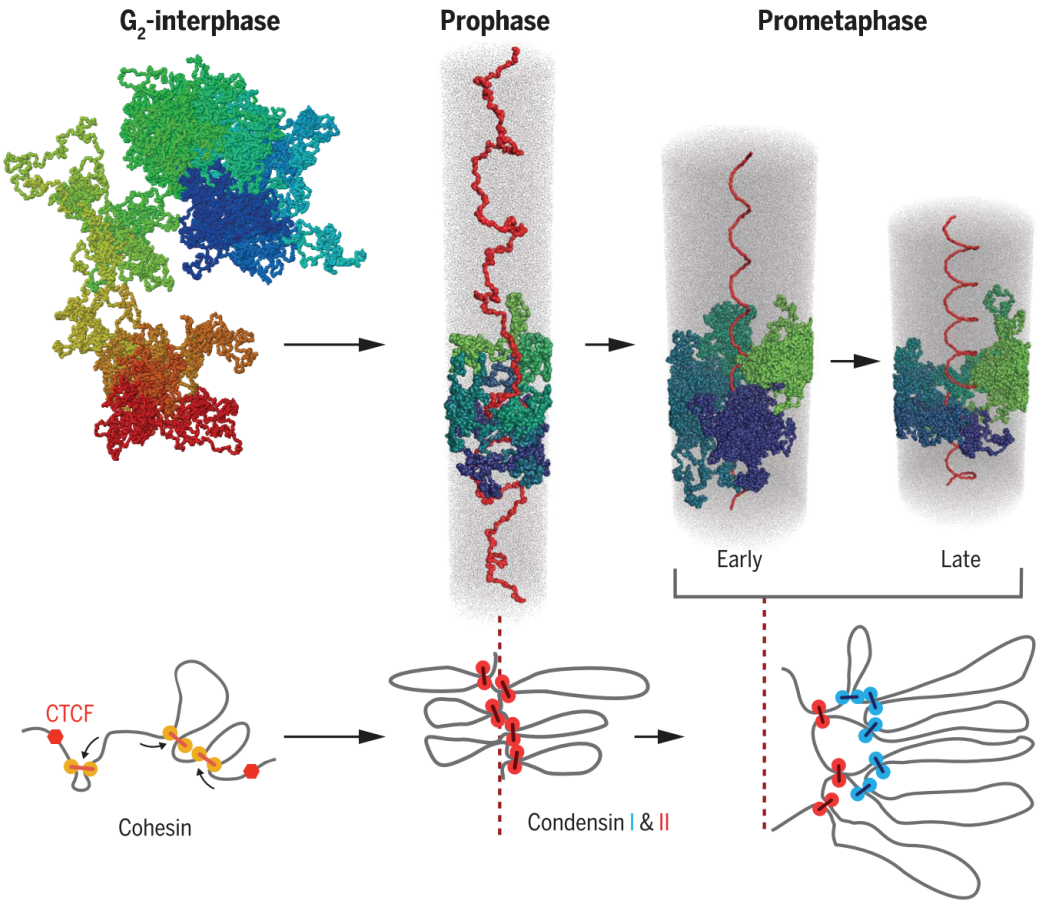


**Figure 1.9 - Integration of the radial loop and loop extrusion models**

Model for the formation of mitotic chromosomes by loop extrusion. In interphase, chromatin is organised into loops by cohesin and CTCF. Condensation begins in prophase by formation of sequential loops by condensin II, leading to a long cylindrical structure. As mitosis proceeds, the condensin II loops grow larger but are further divided by loops-within-loops through condensin I action, maintaining a thin diameter of the chromosome. Chromosomes are also axially shortened by helical coiling of the condensin II scaffold, possibly through the action of other scaffold proteins.
From Gibcus *et al.* 2018.

This updated radial-loop model thus fits much of the data from a variety of approaches. Some factors known to be involved in chromosome structure are not addressed, such as the role of non-condensin scaffold proteins TopoIIα and Kif4 or histone modifications. However, the model provides a good framework within which to consider the remaining mysteries of mitotic chromosome structure to which elements can be added as they become understood.

## 1.3 The F1.1 Cell Line

### 1.3.1 Background of the Cell Line

The cell line which is the primary focus of this study was originally described in Allshire *et al.* 1987 and McManus *et al.* 1994. Named F1.1, the cell line was derived from the mouse mammary tumour cell line C127 through fusion with protoplasts of the fission yeast *Schizosaccharomyces pombe*. A mouse selectable marker, the G418 resistance gene SV2NEO, had previously been inserted at the *ura4* locus on *S. pombe* chromosome III. Fusion products were selected for G418 resistance in order to isolate C127 cells that had taken up *S. pombe* chromatin, at least from the SV2NEO integration and surrounding regions but potentially including a large amount of associated chromatin. Several such fusion cells were isolated, some of which were shown to contain unstable fragments of unintegrated *S. pombe* DNA while others were shown to have stably integrated large stretches of *S. pombe* DNA in the mouse genome (Allshire *et al.* 1987). F1.1 was among the latter.

Studies investigating the fate of large amounts of DNA introduced to mammalian cells have revealed two common features. Firstly, whether the DNA is introduced in small or large blocks, substantial rearrangements of the DNA will occur prior to its integration into the genome (Wigler *et al.* 1978, Perucho *et al.* 1980). These rearrangement events are thought to take place in the cytoplasm as a result of various enzymes degrading, ligating and recombining the foreign DNA (Folger *et al.* 1982, Luthman and Magnusson 1983). Secondly, the actual integration of DNA into the host genome is a rare event and is unlikely to occur more than once or a very few times (Folger *et al.* 1982). The result is that foreign DNA is usually inserted at a single locus in the host genome and is most often present as a single large block of

highly rearranged DNA. However, in most of these early studies DNA was introduced to mammalian cells by transfection, usually by the calcium-phosphate method, while in F1.1 the *S. pombe* DNA was introduced by cell fusion.

Nevertheless, the fate of the *S. pombe* DNA integrated into the F1.1 genome was generally found to conform to the pattern observed for transfected DNA. Analysis of the DNA content of the F1.1 foreign insert found that it contained DNA from substantial portions of *S. pombe* chromosome III, as would be expected given the site of integration of the SV2NEO gene on that chromosome, but also contained sequences from at least parts of the other two *S. pombe* chromosomes I and II, as well as centromeric sequences from all three chromosomes (McManus *et al.* 1994) (**Figure 1.10**). FISH performed on F1.1 chromosome spreads using probes for total *S. pombe* showed only a single large fluorescent signal, approximately two-thirds of the way down a mouse chromosome arm, suggesting that all of the *S. pombe* DNA had been inserted at a single locus in the mouse genome and must have undergone a certain amount of rearrangement prior to its insertion into the genome. On the other hand, Southern analysis using *S. pombe* probes of a few kb showed the target regions to be intact in F1.1, unlike in other mouse cell lines in which the DNA was introduced by transfection. The rearrangement was thus judged to have occurred at a scale above that detectable by these methods, while transfected DNA appeared to have been more substantially scrambled.

The original analyses thus estimated that in F1.1 around 5 to 10 Mb of DNA from *S. pombe*, primarily consisting of sequence from chromosome III but also a smaller amount from chromosomes I and II, had been inserted at a single locus in the C127 mouse genome with some degree of rearrangement, though it remained intact at the kb level.

### 1.3.2 Structure of the Insert

Both light and electron microscopy revealed the *S. pombe* insert to have a unique structure on the mitotic chromosome in F1.1 (McManus *et al.* 1994) (**Figure 1.11**). By light microscopy the width of the chromatid at the site of insertion was considerably narrower than along the rest of the chromosome, measured to average 1.20 µm compared with 1.74 µm. The intensity of the DNA stain was also reduced

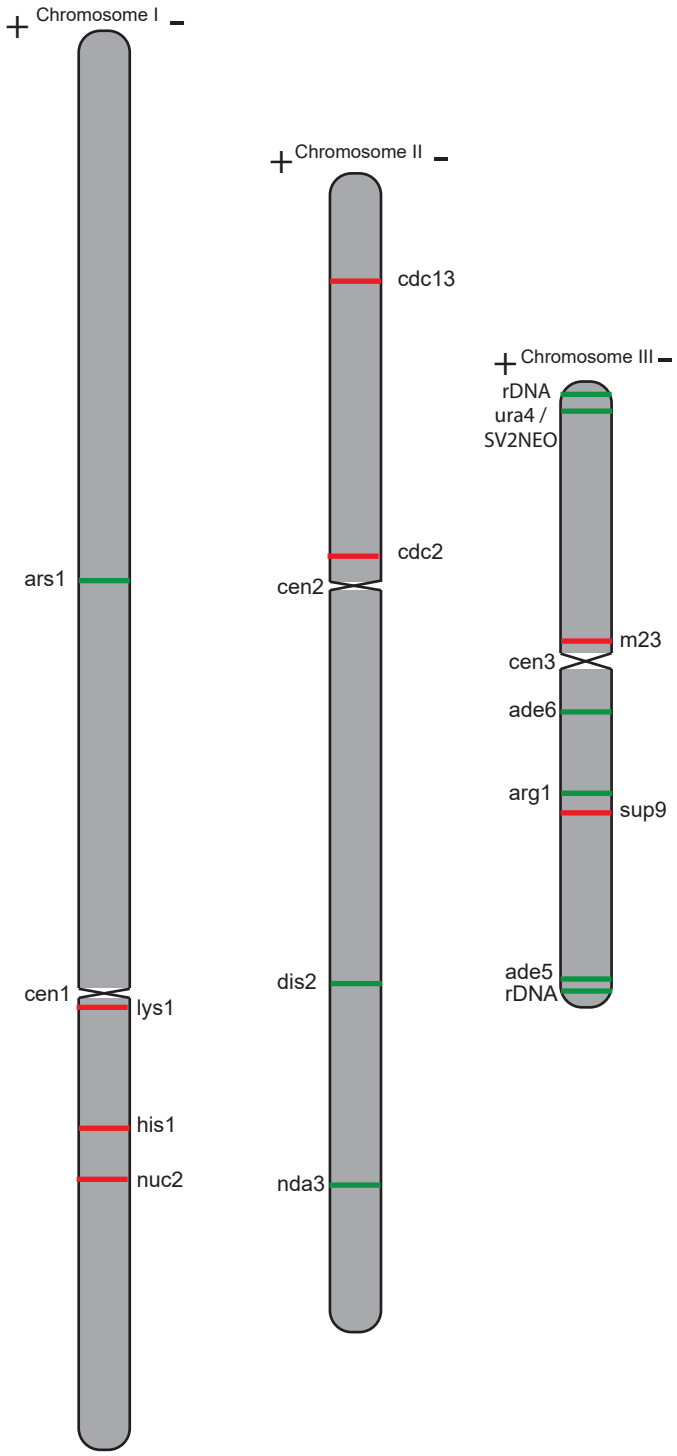


**Figure 1.10 - Investigation of the DNA content of the F1.1 insert region**

Summary of the Southern analysis results from McManus *et al.* 1994. Regions of *S. pombe* DNA probed in F1.1 are highlighted at their location within the *S. pombe* genome. Regions found to be present in F1.1 are highlighted in green and labeled on the left side of the chromosome, regions not detected in F1.1 are highlighted in red and labeled on the right side. Centromere regions are marked as constrictions and were present in F1.1.

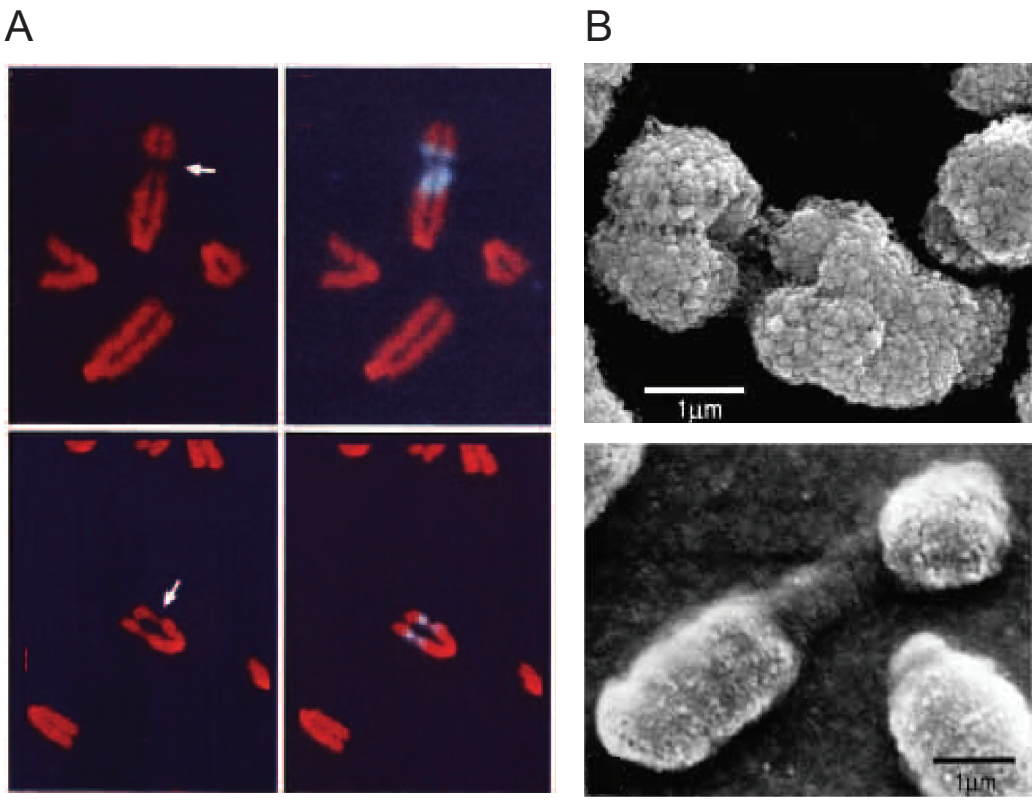


**Figure 1.11 - Distinct structure of the inserted S. pombe chromatin in F1.1**

Observed structure of the foreign *S. pombe* chromatin in the mouse F1.1 cell line by light (**A**) and electron (**B**) microscopy from McManus *et al.* 1994. In each case the insert appears as a region of constricted chromatin. In **A** DNA is stained with propidium iodide and *S. pombe* DNA is highlighted by FISH with probes made from total *S. pombe* DNA. In **B** chromosomes were prepared by the OTOTO procedure and imaged by scanning electron microscopy.

on the foreign chromatin. Electron microscopy confirmed the presence of a chromosome with a region of reduced width at a position consistent with the foreign insert, showing what appeared to be a major constriction of the chromatin on the mitotic chromosome, with separated sister chromatids clearly visible in some cases.

C-banding and silver staining showed no distinctive pattern over the *S. pombe* DNA, in contrast to endogenous mouse centromeres and nucleolar organising regions (NORs) as well as a neocentromere present in the same cell line on another chromosome, which was stained by both techniques (McManus *et al.* 1994). The insert region was therefore distinguished from these other common sites of constriction, although it was suggested that it shared some features in common with fragile sites, which have also been suggested to have abnormal chromatin structures (McManus *et al.* 1994). The inserted *S. pombe* DNA also acquired cytosine methylation over time, but this was not found to correlate with any change in chromatin structure. No particularities were observed with respect to chromatin at the level of nucleosome assembly. It was thus concluded that any alteration of the chromatin structure must be occurring at higher levels of organisation, for which they postulated an increased attachment to the protein scaffold and thus an organisation into smaller loops of chromatin than the surrounding endogenous chromatin.

## 1.4 Aims of the Project

The existence of this highly distinctive chromatin assembled on *S. pombe* DNA inserted in a mouse chromosome raises many questions about the place of structural variation in models of mitotic chromosome folding. As already mentioned, such variation is most obvious at the primary constrictions of centromeres but is also seen at reproducible secondary constrictions that are thought to correlated with certain elements such as NORs and fragile sites (Babu and Verma 1985, Durkin and Glover 2007). While some models of chromatin folding could account for such variation by altering their parameters, for instance by locally changing condensin I and II ratio, altering their residency time and processivity or introducing other structural organisers, the actual existence of these processes and how they might be regulated are not evident. The further study of this variation is also hindered by several complications. The link between secondary constrictions and particular regions or features is not robust and the study of centromeric regions is made

difficult by the nature of their underlying DNA sequence. In many eukaryotes, including vertebrates, centromeres are assembled on large regions of repetitive sequence, making their study or manipulation challenging.

The F1.1 cell lines thus presents a unique opportunity for the study of structural variation along mitotic chromosomes. The distinctive structure was found to be robustly associated with the foreign *S. pombe* chromatin, and investigation of another mouse-*S. pombe* fusion cell line from the same experiment suggested that such an appearance was not unique to F1.1, but a common feature of foreign chromatin in mouse cells (McManus *et al.* 1994). Furthermore, while the precise sequence of the inserted DNA is not yet known, its source from total *S. pombe* DNA suggests that it is primarily composed of unique sequence, making experimentation less complex.

The primary aim of the project is to investigate the nature of the DNA and chromatin structure of the F1.1 insert in more detail through the application of the new techniques that have arisen in the twenty years since the last published work on the cell line. By interpreting these and previous results in the light of new models of mitotic chromosome structure, a clearer understanding of the nature of the structural variation observed in F1.1 and what implications this has for mitotic chromosome structure more generally will be developped.

To achieve that broader goal the specific aims of the project are:

1. Determine the nature of the unusual structure of the chromatin assembled around the F1.1 *S. pombe* DNA insert in mitosis and how it fits with current models of mitotic chromosome structure.

2. Identify what factors are influencing the chromatin structure at the *S. pombe* insert, resulting in a localised alteration in the appearance of the mitotic chromosome.

3. Determine the reason for these factors localising or directing the alteration in chromosome structure to the foreign *S. pombe* DNA specifically.

# Chapter 2: Materials and Methods

## 2.1 Mammalian Cell Protocols

### 2.1.1 Cell Culture

All somatic mouse and human cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) supplemented with 10% Fetal Bovine Serum (FBS, Gibco) and 100U/mL Penecillin-Streptomycin (Gibco). Where cells were drug resistant 400µg/mL G418 or 2µg/mL puromycin (Gibco) was added.

E14 stem cells were grown in plates coated with 0.1% gelatin (Sigma) and maintained in E14 media: GMEM, 15% FBS, 100 U/mL Pen/Strep, 0.1mM NEAA, 1,000 U/mL LIF, 1mM Sodium Pyruvate, 0.1 mM β-mercaptoethanol, 2mM L-Glutamine

All cells were grown in sterile tissue culture flasks or dishes (TPP or Sarstedt). Confluent cells were washed with sterile D-PBS (Gibco) and detached by addition of Trypsin-EDTA (Gibco) for 5 minutes at 37°C. Trypsin was inactivated by the addition of the appropriate medium and split to a new sterile flask either directly after addition of medium or after centrifugation at 300xg for 5 minutes and resuspension in an appropriate volume of medium.

For long-term storage cells were collected as above, centrifuged at 300xg and resuspended in either the appropriate complete medium or FBS. In both cases the freezing medium was supplemented with 10% dimethyl sulfoxide (DMSO, Gibco). Cell suspensions were split into cyrovials and frozen overnight at -80°C before transfer to liquid nitrogen storage. To recover frozen cell lines, samples were thawed at 37°C and immediately transferred to the appropriate medium. Cells were then seeded in flasks either directly or after a centrifugation step at 300xg to remove any excess DMSO.

### 2.1.2 Metaphase Spread Preparation

Metaphase spreads were prepared by one of two methods, depending upon follow up experiments.

Spreads that were directly mounted or used only for FISH were prepared by the standard method: Cells were grown to 70% confluence and collected by mitotic shake off. Occasionally cells were drugged with 0.1µg/mL colcemid for 3 to 4 hours to increase mitotic index, however unless specified this step was omitted. Cells were centrifuged at 800xg for 8 minutes and resuspended drop by drop in 5 to 10mL 100mM KCl pre-warmed to 37°C, mixing continuously. Cells were then incubated at 37°C for 15 minutes, centrifuged again and resuspended in 5 to 10mL fixing solution (3:1 mix of methanol: acetic acid) at -20°C. Cells were centrifuged and washed in fixing solution in the same way twice more and finally resuspended in an appropriate volume of fixing solution (depending on cell number). 10µL of fixed cell suspension was dropped onto a glass slide from approximately 0.5m and allowed to dry before mounting in Vectashield. Slides to be used for FISH were allowed to dry for 2 to 5 days before proceeding.

Spreads that were used in immunolocalisation or immunoFISH experiments were prepared using a Cytospin. Cells were grown and collected by mitotic shake off as above. After centrifugation an 800xg for 5 minutes cells were resuspended in 1mL D-PBS and transferred to a 1.5mL tube. Cells were counted on a haemocytometer to determine cell number and, after centrifugation again, were resuspended at $3x10^4$ cells/mL in 100mM KCl pre-warmed to 37°C. Cell suspensions were then incubated for 15 minutes at 37°C and 100µL of cells were spun onto glass slides using a Cytospin centrifuge at 1,800 rpm for 10 minutes. The slides were then incubated in a coplin jar for 10 minutes at room temperature in KCM buffer before proceeding to immunolocalisation or fixing.

KCM buffer: 120mM KCl, 20mM NaCl, 10mM Tris-HCl pH 8.0, 0.5mM EDTA, 0.1% Triton X-100

### 2.1.3 Transfection

Mammalian cells were usually transfected either using Fugene HD transfection reagent (Promega) or by Neon transfection (Thermo Fischer Scientific). For the transfection of large amounts of *S. pombe* DNA, the calcium phosphate method was used.

For Fugene transfection, the transfection reagent and DNA solutions were first allowed to warm to room temperature. DNA and Fugene were mixed in serum-free OPTIMEM (Gibco) at a ratio of 3:1 as per manufacturer's instructions (generally mixing 2µg DNA with 6µL Fugene in OPTIMEM to a final volume of 100µL, or equivalently scaled). The Fugene-DNA mixture was then incubated for 15 minutes at room temperature. Cultured cell vessels were washed once in D-PBS before addition of an appropriate volume of OPTIMEM. Fugene-DNA was then added dropwise to the cells.

Neon transfection was also performed on F1.1 cells as per the manufacturer's instructions. Cells were trypsinised and collected as normal. Cells were centrifuged at 500xg for 5 minutes, washed in D-PBS and centrifuged again, removing all residual D-PBS. Cells were then resuspended in R-buffer at a concentration of $1x10^7$ cells/ mL. DNA was then added to the suspension to make up a maximum of 10% of the total volume, generally for a final concentration of 0.1µg/ µL DNA. For electroporation, 3mL of electrolytic buffer was added to the Neon tube and inserted into the Neon pipette station. Either 10 or 100 µL of cell suspension were then pipetted into a Neon pipette tip, inserted into the Neon tube and electroporated at 1400V for 20ms and two pulses. Cells were then transferred immediately to an appropriate volume of cell culture medium and plated in an appropriately sized vessel.

F1.1 cell lines stably expressing PREditOR constructs were obtained by co-transfecting a plasmid expressing the Sleeping Beauty transposase with the self-designed plasmids including the tet-inducible constructs and a puromycin resistance gene within a Sleeping Beauty transposon. Cells with the transposon inserted were assessed by puromycin treatment.

For Calcium Phosphate transfection cells were grown to 30-40% confluence. 15-20µg DNA was added to 124µL of 2M $CaCl_2$ and the total volume brought to 1mL. 1mL 2xHBS was then added drop-wise while aerating constantly. 1mL of the final mix was added dropwise to a 10cm diameter plate, or relatively equivalent volumes to other sized plates.

2xHBS, pH 7.05: 50mM HEPES pH 7.05, 10mM KCl, 12mM dextrose, 280mM NaCl, 1.5mM $Na_2PO_4$

### 2.1.4 Immunolocalisation

Immunolocalisation was performed slightly differently depending on the nature of the slides or coverslips.

For interphase cells grown directly on slides or coverslips, cells were first fixed in 4% PFA (paraformaldehyde), 1xPBS for 20 minutes at 37°C. Slides were then washed briefly in 1xPBS followed by permeabilisation in permeabilisation solution (1xPBS, 0.2% Triton X-100) for 2 minutes. Slides were then blocked in blocking solution (1% BSA in permeabilisation solution) for 1 hour at 37°C. Slides or coverslips were then placed on 100 or 50µL (respectively) of primary antibody diluted in blocking solution on parafilm in a humidity chamber for 1 hour at 37°C, washed 3 times for 10 minutes in permeabilisation solution, incubated similarly with secondary antibody for 45 minutes at 37°C and washed twice in permeabilisation solution and once in 1xPBS for 10 minutes each before mounting and sealing.

For metaphase spreads prepared by cytospin, following the KCM incubation slides were placed on 100µL blocking solution (1% BSA in KCM buffer) on parafilm in a humidity chamber as above for 30 minutes at 37°C followed by the primary then secondary antibodies, each diluted in blocking solution, for 30 minutes and 45 minutes respectively at 37°C. After each antibody incubation the slides were washed twice in KCM buffer for 5 minutes. The slides were then fixed in 4% PFA in KCM for 10 minutes at 37°C and washed briefly once in KCM and twice in water before mounting and sealing.

Antibodies used were as follows:

Primary antibodies: Rabbit α-H3K9me3 (active motif 39161), rabbit α-SMC2 and α-SMC3 (from Ana Losada), rabbit α-H4 (Sigma SAB4500312), rabbit α-CAPH1 (Abcam Ab154105)

Secondary antibody: Donkey α-rabbit Alexa594 (Invitrogen A21207)

### 2.1.5 FISH

Slides for FISH were prepared as follows. If the slides had been previously fixed in PFA (whether interphase cells or cytospun metaphase spreads) they were first washed for 2 minutes in permeablisation solution (for interphase cells) or KCM buffer (for cytospun spreads) followed by 2 minutes in 2xSSC. These washes were omitted for regular metaphase spread preparations.

All types of slide were then incubated in 2xSSC with 100µg/mL RNase at 37°C for 1 hour. Cells were then briefly washed in 2xSSC followed by dehydration through a series of 2 minute washes in 70%, 90% and 100% ethanol and air dried. Slides were then denatured in 2xSSC, 70% formamide in different conditions: pH 7.5 for 1.5-2 minutes at 70°C for regular metaphase spreads, pH 7.2 for 20 minutes at 80°C for PFA fixed slides. Denaturation was rapidly stopped by dipping slides into 70% ethanol on ice, followed again by dehydration through 90% and 100% ethanol at room temperature and air drying.

Meanwhile 100-200ng of previously labelled FISH probes or appropriate amounts of chromosome paint (as per manufacturer's instructions) per slide were mixed with 5µg sonicated salmon sperm DNA and 2 volumes of 100% ethanol. For probes to *S. pombe* DNA this was sufficient, however if the probes were chromosome paints or for other regions of mouse DNA 10µg mouse Cot1 DNA was also added, with the appropriate increase in ethanol added. Probes were evaporated on a PCR block at 65°C. Probes were then allowed to re-dissolve for 1 hour in 15µL hybridisation mix (2xSSC, 50% deionised formamide, 10% dextran sulfate, 1% Tween 20).

Once the slides were ready, probes were thoroughly mixed and denatured at 70°C for 5 minutes and either snap-cooled on ice or (if Cot1 was included) allowed to anneal for 15 minutes at 37°C. 15µL of probe were then added to a 22x22mm coverslip, placed on the slide and sealed with rubber cement to incubate overnight at 37°C in a humidity chamber.

The following day slides were washed 4 times for 3 minutes in 2xSSC at 45°C and 4 times for 3 minutes in 0.1xSSC at 60°C before a brief wash in 4xSSC, 0.1% tween 20 at 37°C. Blocking was performed for 5 minutes in in 4xSSC, 5% milk powder. Slides were then incubated with the appropriate detection antibodies for 30 to 60 minutes in 4xSSC, 5% milk powder under a coverslip. After each antibody incubation slides were washed 3 times for 2 minutes in 4xSSC, 0.1% tween 20 at 37°C. After all antibody incubations slides were mounted directly in mounting medium and sealed.

Except where indicated all antibodies were from Vector Labs and were used as follows:

For detection of biotinylated probes: sequential incubation of Avidin-FITC (1:500), biotinylated α-Avidin (1:100) and Avidin-FITC again.

For detection of DIG-labeled probes: Rhodamine α-DIG Fab fragments (Roche 1:20) followed by Texas Red α-sheep antibody (1:100).

### 2.1.6 iPS reprogramming

Transfection based iPS reprogramming was performed after the method in Kaji *et al.* 2009. Briefly, the MKOS plasmid expressing the four reprogramming factors was co-transfected with a plasmid expressing the *PiggyBack* transposase to insert MKOS plasmid into the genome. A reporter plasmid expressing GFP was also included to monitor cells that had taken up DNA. 2 days post-transfection the cell media was switched to ES media and a further 5 days later the media was switched to knock-out media to stimulate reprogramming.

ES media: DMEM, 10% FBS, 100U/mL Pen/Strep, 0.1mM NEAA, 1,000 U/mL LIF, 1mM Sodium Pyruvate, 50 μM β-mercaptoethanol, 2mM L-Glutamine

Knockout media: DMEM, 15% knockout serum, 100U/mL Pen/Strep, 0.1mM NEAA, 1,000 U/mL LIF, 1mM Sodium Pyruvate, 50 μM β-mercaptoethanol, 2mM L-Glutamine

Lentiviral reprogramming used the Merck-Millipore STEMCCA Cre-excisable dox-inducible polycistronic OKSM lentivirus reprogramming kit (catalogue number SCR513) and followed manufacturer's instructions.

### 2.1.7 Cell fusion

Mammalian cells were fused with PEG as follows. All centrifugation steps were performed at 300xg for 5 minutes. Cells were trypsinised as normal, collected and centrifuged, then washed with serum-free DMEM (SF-DMEM, i.e. without FBS) and centrifuged again. $5x10^7$ cells of each cell type were combined in a tube in SF-DMEM, mixed and centrifuged together. The mixed cell pellet was then resuspended slowly in 1mL of 50% PEG 1500 (Roche) at 37°C over the course of 1 minute and incubated for a further 1 minute at 37°C. The total volume was brought to 5mL with SF-DMEM gradually over 4 minutes, and then to 15mL adding SF-DMEM drop by drop. Cells were then incubated for a further 5 minutes at 37°C before centrifugation and washing twice in SF-DMEM. Cells were finally resuspended in the appropriate cell culture media and plated.

For fusion of mammalian cells with *S. pombe* protoplasts, protoplasts were prepared by Alison Pidoux as described in **Section 2.6.2**. $1x10^8$ protoplasts per fusion were pelleted at 1600xg for 3 min. Mammalian cells were prepared as indicated above. $1x10^7$ cells in 5mL SF-DMEM were then added to the tube containing the *S. pombe* protoplast pellet, being careful not to dislodge the pellet. Mammalian cells were pelleted on top of the *S. pombe* cells by centrifugation at 300xg for 5 minutes. PEG fusion was then carried out just as described above for mammalian cells.

## 2.2 Microscopy

Most microscope images were acquired on a DeltaVision Elite system (Applied Precision) with an Olympus UPlanSApo ×100 oil immersion objective and a Lumencor SpectraX LED light source. Camera (Photometrics Cool Snap HQ), shutter and stage were controlled through SoftWorx (Applied Precision). Other images were acquired on a Nikon Eclipse Ti2 using the NIS-Elements software for instrument control and image acquisition.

Z-series were collected and subsequently projected into a single image using either SoftWorx or ImageJ software (National Institutes of Health, Bethesda, MD). For intensity measurements DeltaVision images were also deconvolved in Softworx. Intensity and chromosome length measurements were taken using ImageJ. Intensity measurements were then binned as a function of their relative position along the chromosome and averaged. For statistical analysis, the region of interest (*S. pombe* DNA) was defined by the 3, 4 or 5 points along the chromosome with the highest average FISH values (or lowest for chromosome 10 paint), depending on the cell line or experiment. This region was then compared to a region of similar size along the same chromosome, excluding the centromere region of 0-25% of the chromosome length and the telomere region of 95-100%. Comparison was done by the Kolmogorov Smirnoff test (KS test). Significance threshold was set at 0.01, with a critical D of $1.63/\sqrt{n}$, where n is the number of measurements.

## 2.3 DNA Protocols

### 2.3.1 DNA extraction

Mammalian Cells were trypsinised and centrifuged as usual and washed twice in ice-cold D-PBS. Cells were then resuspended in 0.3mL of digestion buffer. The samples were incubated overnight at 50°C shaking lightly and DNA was extracted with an equal volume of Phenol:Chloroform and centrifuging at 17,000xg for 10 minutes. The DNA in the aqueous phase was collected and precipitated briefly with a 1/10 volume of NaOAc and 2 volumes of ethanol, centrifuging immediately at 17,000xg for 2 minutes and washing once in 70% ethanol. After drying briefly, the DNA was resuspended in water or TE buffer.

Digestion buffer:
100mM NaCl
10mM Tris pH 8
25mM EDTA pH 8
0.5% SDS
0.1mg/mL Proteinase K

### 2.3.2 PCR

PCR reactions were performed either using Taq Polymerase (Roche), Pfx Polymerase (Life Technologies) or Hot Start Taq (NEB) as follows:

Taq reaction (20µL):
1xTaq Buffer
100µM dNTPs
0.5U Taq Polymerase
1µM each primer
Template DNA
$dH_2O$

Taq PCR programme:
1. 94°C for 4 minutes
2. 94°C for 30 seconds
55°C for 30 seconds
72°C for 1 minute per kb of product
Repeat x 34 cycles
3. 72°C for 5 minutes
4. 4°C for 5 minutes

Pfx reaction (20µL):
1xPfx buffer
250µM dNTPs
0.02U Pfx Polymerase
1mM $MgSO_4$

0.5µM each primer
DNA Template
$dH_2O$

Pfx PCR programme:
1. 94°C for 2 minutes
2. 94°C for 15 seconds
52°C for 1 minute
68°C for 1 minute per kb of product
Repeat x 34 cycles
3. 68°C for 10 minutes
4. 4°C for 2 minutes

Hot Start Taq reaction (20µL):
1xTaq buffer
200µM dNTPs
0.5U Hot Start Taq Polymerase
1µM each primer
Template DNA
$dH_2O$

Hot Start Taq programme:
1. 95°C for 30 seconds
2. 95°C for 30 seconds
55°C for 30 seconds
68°C for 2.5 minutes
Repeat x30 cycles
3. 68°C for 5 minutes
4. 4°C for 3 minutes

Template DNA used was approximately 1ng for plasmid DNA, 10ng for *S. pombe* genomic DNA and 100ng for mammalian genomic DNA.

### 2.3.3 Agarose Gel Electrophoresis (AGE)

Agarose was dissolved in 1xTBE (1M Tris, 1M boric acid, 20mM EDTA) at a final concentration between 1 and 2% w/v depending on the size of the DNA fragments to be analysed. Ethidium bromide was added at a concentration of 0.2µg/mL before solidification of the gel in a mould. 6xLoading dye (NEB) was added to DNA samples before loading. After migration, DNA was visualised by UV using a transilluminator.

If required, DNA was gel extracted using the Qiagen Gel Extraction kit.

### 2.3.4 Quantitative PCR (qPCR)

qPCR reactions were performed in 96 or 384 well plates with 10µL of PCR reaction per well.

qPCR reaction (10µL):
5µL Light Cycler 480 SYBRGreen Master Mix (Roche)
0.05µL of each primer (100µM),
1.9µL of $dH_2O$
3µL DNA sample.

qPCR programme:
Initial Denature at 95°C for 2 minutes
44 to 55 cycles:
95°C for 20 seconds
55°C for 20 seconds
72°C for 20 seconds

The results were analysed with the Light Cycler 480 Software (Roche).

### 2.3.5 Sanger sequencing

Sanger sequencing was performed using the BigDye Terminator Cycle sequencing kit.

Sequencing reaction (20µL):
2µL BigDye Terminator reaction mix
0.25µM each primer
100ng-1µg template DNA
$dH_2O$

Sequencing program:
1. 96°C for 1 minute
2. 96°C for 10 seconds
50°C for 5 seconds
60°C for 4 minutes
Repeat x 25 cycles

Analysis of the sequence was then done by Edinburgh Genomics.

### 2.3.6 Nick Translation

Probes for FISH were labelled using nick translation on genomic *S. pombe* DNA.

Nick translation reaction (20µL):
1 x Nick translation salts
62.5µM dATP
62.5µM dCTP
62.5µM dGTP
125µM bio-16-dUTP (Roche) or 25µM dTTP + 75µM digoxigenin-11-dUTP
1 – 5 µg DNA
1U DNase1 (Roche)
3U DNA polymerase I (Invitrogen)
$dH_2O$

The reaction was incubated for 90 minutes at 16°C and inactivated by addition of SDS to 1% and EDTA to 60mM. Samples were made up to 90µL with TE (10mM Tris pH 8, 1mM EDTA) and purified by centrifugation through a Roche Sephadex G-50 column.

Probe concentrations were determined by spotting on a nitrocellulose gridded circle filter (Whatman) along with previously made standards from the lab of Wendy Bickmore. Samples were crosslinked onto the filter by exposure to UV light at 1.5 x $10^5$ µJ/$CM^2$. The filter was washed briefly in NT buffer and then incubated in NT buffer with 3% milk powder for 15-30 minutes at 60°C. Filters were then labelled with either 1U/mL streptavidin alkaline phosphatase (Roche) for biotin labelled probes or 0.75U/mL anti-dig alkaline phosphatase (roche) for dig-labelled probes, in NT buffer for 15 minutes at room temperature. Filters were washed twice in NT buffer for 15 minutes followed by a brief wash in 0.1M Tris pH 9.5. The labelled spots were then visualised using the BCIP/NBT kit from Vector Labs, incubating the filter in Tris pH 9.5 containing one drop of each kit reagent per 2.5mL of buffer. Visible spots were compared to standards to determine the concentration of each probe.

10 x NT salts:
0.5M Tris pH 7.5
0.1M $MgSO_4$
1mM DTT
0.5mg/mL BSA fraction V

NT buffer:
0.1M Tris pH 7.5
0.15M NaCl

### 2.3.7 Bisulfite conversion

Bisulfite conversion of mammalian genomic DNA was performed using the Qiagen EpiTect Bisulfite kit and followed the manufacturer's instructions. Selected targets were then amplified by Hot Start PCR using primers designed for converted DNA, with targets of 200-300bp. PCR products were then subjected to AGE, gel extracted and cloned into a vector using the Agilent Strataclone PCR Cloning kit. Cloned targets were then sequenced as indicated above.

### 2.3.8 Whole genome sequencing

Whole genome library preparation and sequencing was performed on F1.1 genomic DNA by BGI Tech on an Illumina HiSeq 2500, generating paired-end 125bp reads. Sequence assembly was performed by Pin Tong and mapped to *Mus musculus* or *Schizosaccharomyces pombe* reference genomes.

### 2.3.9 Hi-C

$2.5x10^7$ asynchronous and mitotic F1.1 cells were collected to generate the two Hi-C datasets. Asynchronous cells were grown to near confluency and then fixed in their culture plate with 1% paraformaldehyde (PFA, Sigma) for 10 minutes at room temperature. Fixation was stopped by addition of glycine to a final concentration of 125 mM for 5 minutes at room temperature and 15 minutes on ice. Cells were then collected by scraping and centrifuged at 800xg for 10 minutes. The liquid phase was completely removed by aspiration and the cell pellet was snap frozen in liquid nitrogen and stored at -80°C.

To obtain mitotic cells a thymidine block was performed. Cells around 70% confluency were treated with 2mM thymidine for 18 hours. Cells were then washed twice with D-PBS and given regular cell culture medium. Mitotic cells were collected by shake-off twice after 5 hours and 8 hours growth in normal media, respectively. After each collection, cells were centrifuged at 300xg for 10 minutes and resuspended in SF-DMEM. Cells were then fixed and frozen as above.

Preparation of the Hi-C library by HindIII digestion, sequencing and analysis was performed by the lab of Job Dekker.

## 2.4 RNA Protocols

### RNA Extraction and RT-qPCR

Total RNA was extracted from mammalian cells using the RNeasy Mini Kit (Qiagen) as per the manufacturer's instructions. 15µg RNA was then DNase treated with 2µL Ambion DNase (Life Tech.) in a 50µL reaction for two successive 1 hour steps at

37°C, adding fresh DNase each time. The RNA was then purified using the RNeasy Mini Kit.

The RNA was then reverse transcribed into cDNA as follows. 1µg RNA was added to 1µL Random Hexamers (100ng/µL) and 1µL dNTPs (10mM), bringing the total volume to 13µL with $H_2O$. Two identical such samples were prepared for each RNA sample. The samples were denatured at 65°C for 5 min and immediately put on ice to prevent re-annealing. To one sample was added 4µL 5x buffer, 1µL DTT, 1µL Reverse Transcriptase (RT) and 1µL $H_2O$. The other sample was made a no RT control by omitting the RT and substituting a further 1µL of $H_2O$.

Samples were then run on a PCR block for:
25°C for 5 minutes
50°C for 1 hour
70°C for 15 minutes

The samples were then diluted 1:20 before qPCR as described in **Section 2.3.4**. Analysis was as normal, with normalisation being performed relative to a strongly expressed positive control.

## 2.5 Protein Protocols

### Chromatin Immuno-Precipitation (ChIP)

Mammalian cells were grown to near confluence in tissue culture dishes before fixation in growth medium with 1% PFA for 10 minutes at room temperature. Fixation was stopped by addition of glycine to a final concentration of 125 mM for 5 minutes at room temperature before washing twice in ice cold PBS. Cells were then harvested in PBS by scraping and spun down at 800xg for 5 minutes at 4°C. The fixed cells were then washed once each in 5mL Wash Buffer 1 and Wash Buffer 2, with a 10 minute incubation step on ice after each resuspension.

Cells were then resuspended in between 0.3 and 0.6 mL of Lysis Buffer and stored at 4°C until sonication up to a maximum period of 24h. Cells were then sonicated on a Diagenode Bioruptor for 15 minutes on a 30 second on/off cycle to obtain

fragments between 400 and 600 bp. The sonicated chromatin samples were then centrifuged at 17,000xg for 10 minutes and 4°C, collecting the supernatant to clarify the lysate.

The DNA concentration of the samples was measured on a Nanodrop and separated into aliquots with the appropriate amount of chromatin. 25 to 50 µg of were used to IP histone modifications (generally 40µg) and 100µg for other chromatin associated proteins. These aliquots were made up to 100µL with Lysis Buffer before addition of 900µL of ChIP Dilution Buffer. 4µg of the appropriate antibody was then added to the samples, including a no-antibody control for each ChIP, which were then shaken gently overnight at 4°C. 10µg were taken at each ChIP as an input, made up to 50µL with Lysis Buffer and stored at 4°C for later use.

50µL of pre-washed protein G Dynabeads (Thermo Fisher) were aliquoted for each sample, and the supernatant removed with the aid of a magnetic rack. The samples were then added to the beads and shaken again at 4°C for 4 to 5 hours. The samples were then washed sequentially once in RIPA, High Salt Wash and LiCl Wash and twice in TE Buffer: in each case removing the supernatant with the magnetic rack, adding 900µL of the wash, vortexing briefly and shaking for 10 minutes at 4°C.

DNA was then eluted from the Dynabeads by addition of 400µL Elution Buffer and shaking strongly at room temperature. For this step and all subsequent steps the input was included with the IP samples. The eluate was then collected on the magnetic rack and the beads discarded. The DNA-protein crosslinks were then reversed by addition of 20µL 4M NaCl and shaking overnight at 300rpm and 65°C. Proteins were then digested in 0.05mg/mL proteinase K for 1 hour at 55°C, with addition of 8 µL 0.5M EDTA, 16µL 1M Tris pH6.5. DNA was then Phenol-Chloroform extracted and ethanol precipitated before RNase treatment in 0.25µg/mL DNase-free RNAse for 30 minutes at room temperature.

DNA amounts were then determined by qPCR as described in **Section 2.3.4**. Inputs were diluted 1:20 before qPCR. Enrichments were calculated as percentage of DNA immunoprecipitated relative to input at the locus in question, and usually normalised

to a positive control locus. Enrichments were compared by a one-tailed Student's *t*-test with a significance threshold set at 0.05.

Antibodies used were as follows:

α-H3K9me3 (active motif 39161), α-H3K4me3 (active motif 39159), α-H3K36me3 (active motif 61101), α-H3K9ac (active motif 39137), α-H3K27ac (Abcam ab4729), α-HP1α (Abcam Ab77256)

Solutions used were as follows:

Wash Buffer 1: 0.25% Triton X-100, 10mM EDTA, 0.5mM EGTA, 10mM HEPES

Wash Buffer 2: 200mM NaCl, 10mM EDTA, 0.5mM EGTA, 10mM HEPES

Lysis Buffer: 1% SDS, 10mM EDTA, 50mM Tris pH 8.1, 0.1mM PMSF, 1x Protease Inhibitor (Roche)

Dilution Buffer: 167mM NaCl, 16.7mM Tris pH 8.1, 1.2mM EDTA, 1.1% Triton X-100, 0.01% SDS

RIPA: 150mM NaCl, 50mM Tris pH 8, 0.1% SDS, 0.5% Na deoxycholate, 1% NP40

High Salt Wash: 0.5M NaCl, 50mM Tris pH 8, 0.1% SDS, 1% NP40

LiCl Wash: 250mM LiCl, 50mM Tris pH 8, 0.5% Na deoxycholate, 1% NP40

## 2.6 *S. pombe* protocols

### 2.6.1 *S. pombe* growth and manipulation

*S. pombe* culture was performed by either Alison Pidoux or Sharon White following standard culture methods as described (Moreno, 1991). Cells were grown in YES at 32°C. High molecular weight genomic DNA was prepared from *S. pombe* strain 43

using a Qiagen Blood and Cell Culture DNA Kit (Qiagen), according to manufacturer's instructions.

Small scale genomic DNA for FISH probes was prepared as follows. Cells were grown to early stationary phase in 5mL YES. Cells were centrifuged at 1600xg for 2 minutes and resuspended in 0.25mL SP buffer containing 0.4 mg/ml Zymolyase-100T (MP biomedicals) and incubated at 37$^{o}$C for 30-60 min. Cells were centrifuged again and resuspended in 0.5 ml TE buffer. 50µL of 10% SDS was added, mixing well, followed by 165µL of 5M KOAc. Cells were incubated on ice for 30 minutes. Samples were centrifuged at 17,000xg for 10 minutes and the supernatant was transferred to a new tube containing 0.75mL isopropanol. Samples were incubated on dry ice for 10 minutes and DNA was precipitated by centrifuging again at 17,000xg for 10 minutes. DNA was resuspended in 0.3mL TE with 10µg/mL DNase-free RNase Roche) for 30-60 minutes at 37°C. DNA was phenol/chloroform extracted, ethanol precipitated and resuspended in 20µL TE.

SP buffer:
1.2M sorbitol
50mM sodium citrate, 50mM $Na_2HPO_4.7H_2O$
40mM EDTA
pH adjusted to 5.6 with orthophosphoric acid

### 2.6.2 Preparation of *S. pombe* Protoplasts

Protoplasting of *S. pombe* was performed by Alison Pidoux following a procedure modified from previously reported methods (Allshire *et al.* 1987, Flor-Parra *et al.* 2014). Briefly, S. pombe cells were harvested from log-phase cultures grown in YES. 1 x $10^9$ cells were resuspended in 10mL SP2 buffer containing 25mg/ml Lallzyme (Lallemand, SA, supplied by Litmus Wines), then incubated at 36°C for approximately 45 minutes until 80-90% of cells were converted to spherical protoplasts (determined by phase contrast microscopy). Protoplasts were pelleted at 1600xg for 3 min, then gently resuspended in 5mL wash buffer (1.2M Sorbitol, 10mM Tris-HCl pH 7.5) and re-pelleted, a total of 3 times. Protoplasts were counted using a haemocytometer and 1 x $10^8$ cells were aliquoted per fusion. Yield of protoplasts was approximately 10%.

SP2 Buffer:
50mM citrate-phosphate (7.1g/L $Na_2HPO_4$, 11.5g/L citric acid)
1.2M Sorbitol
pH to 5.6 with NaOH

### 2.6.3 *S. pombe* media and strain

YES liquid in 900mL:
5g Yeast extact
30g D-glucose anhydrous
0.2g Adenine
0.2g Arginine
0.2g Lysine
0.2g Histidine
0.2g Uracil
0.2g Leucine

*S. pombe* strain FY43: *h-Int5(pUraSV2Neo) ade6-210 leu1-32 ura4-D18*

## 2.7 Bacterial Protocols

### 2.7.1 Transformation

10 to 50ng of plasmid was transformed in 30µL DH5α competent bacterial cells (Invitrogen). DNA and cells were mixed and incubated on ice for 30 minutes, heatshocked for 45 seconds at 42°C and left on ice for 2 minutes before addition of 1mL LB. Cells were grown for 45 minutes at 37°C and plated onto LB agar plates with the appropriate antibiotic.

The pYIP-CMV-Suv39H9DSET-EYFP plasmid required ccd resistant bacteria and so were transformed into ccdB Survival 2 T1 competent cells (Invitrogen).

### 2.7.2 Plasmid extraction

A single colony of bacteria containing the plasmid of interest was grown in 5mL or 30mL LB with antibiotic at 37°C overnight. Cells were collected by centrifugation at 6000xg for 5-15 minutes and plasmids were extracted using either Qiagen miniprep / midiprep kits or Invitrogen Purelink HiPure midikit.

### 2.7.3 Bacterial cloning

PREditOR plasmids were obtained from the lab of William Earnshaw and included: pYIP-CMV-Suv39H1DSET-EYFP, pYIP-CMV-Suv39H1DSET-EYFP-JMJD2D and pYIP-CMV-Suv39H1DSETmut-EYFP-JMJD2D. The open reading frames from each plasmid were then cloned into the pSBTet-BP plasmid (Addgene #60496) by PCR based cloning.

The regions of interest from each plasmid were PCR amplified to create products with SfiI digestible ends that would align with the cut pSBTet-BP vector. Products were purified using the Qiagen PCR purification kit. Purified products and pSBTet-BP plasmids were then digested with SfiI at 1 U/µL in 1x CutSmart buffer (NEB) at 50°C for 1 hour. The digested vector was dephosphorylated by addition of 1U of calf intestinal phosphatase and incubation for 1 hour at 37°C. All digest products were then gel purified. PCR products and cut vectors were then ligated as follows:

Quick ligation reaction (20µL):
1x Quick ligase reaction buffer
1µL Quick ligase
0.02 pmol vector DNA
0.06 pmol insert DNA
$dH_2O$

The reaction was incubated at room temperature for 5 minutes, after which 2-5µL was transformed into competent bacterial cells.

Other plasmids used were the Sleeping Beauty transposase expressing pCMV(CAT)T7-SB100 (Addgene #34879) and the H2B-GFP expressing plasmid (Addgene #11680).

### 2.7.4 Bacterial Media

LB medium: 10% NaCl, 10% Bacto tryptone, 5% Bacto yeast extract

LB agar: 10% NaCl, 10% Bacto tryptone, 5% Bacto yeast extract, 15% Bacto agar

Supplement: Carbenicillin 50µg/mL, Kanamycin 50µg/mL

## 2.8 Primers

| Primer name | Primer sequence |
|---|---|
| **Sequencing verification:** | |
| dpb10_F | GCCGACTTCTCGTCAAACTC |
| dpb10_R | CTTGGTCCAAAGACCCGTAA |
| rpn6_F | GCCCTCACGAATACCCAGTA |
| rpn6_R | CGTTCGGAAACGATTTGATT |
| pep7_F | GAATGGGGATTCATCAGGTG |
| pep7_R | AAAAGAAGCGCATCAAGCAT |
| saf3_F | GGCAAAAGCCAATGAAGAAG |
| saf3_R | TCGCAAAAAGATCTCCCCTA |
| fan1_F | GGTGTGTGGAAAAAGCAGGT |
| fan1_R | GCGGCAGTGTTTTCAAAGAT |
| isp4_F | GAATTTCGCATTGGACGTTT |
| isp4_R | ACGAATCCAAGTGACCCAAG |
| pof2_F | TATTTGCGCAAGCTGAATTG |
| pof2_R | TCGATTTCTGGAGACCAACC |
| raf2_F | TTGCAGCGTTATTGCTCAAC |
| raf2_R | GCAGGCCAATAACCAAAAGA |
| pog1_F | CACAAGAGGTCCCAGTGGTT |
| pog1_R | CGATCACGAGCACTACTCCA |
| dal51_F | TTCCCATGAGAGGTTGGAAG |
| dal51_R | TGTGGTGATGGTAGCCAAAA |
| mod21_F | GGTTGATGCCTGACTGGTCT |
| mod21_R | CCGTAATCGGCAAGTTCTGT |
| rad31_F | TTGGGCGAAACACTAAAACC |
| rad31_R | ATTCGGCATCAAAAATCCAG |
| arg3_F | TCTCCAGGCGTTAGCAGATT |
| arg3_R | GGATGACGGGGTAAACAATG |
| pfl2_F | GTTCCTCTCAGGGCAGTGTC |
| pfl2_R | GATATCCTGAGCCACCTGGA |
| rpb10_F | CCAACGAGGACTGATCCACT |

| | |
|---|---|
| rpb10_R | AGGCTTCTCGGCCTAATCTC |
| cta3_F | CAGTTTCCATGTGACCAACG |
| cta3_R | TCGCCTTTTACAACCCACTC |
| pho7_F | TTGCGCAAAGTGTCAAAAAG |
| pho7_R | TGTTAGATACCGCCACACCA |
| cut2_F | CATTTGGACCCCATTCAATC |
| cut2_R | TTGCCAGATGCAGTTACTCG |
| rap1_F | AGTGATCCGCGAACCAATAG |
| rap1_R | AGCTTTCTGGAGACCGATCA |
| dad4_F | TGTCCTACCTTCTTGCCATTG |
| dad4_R | GGAAAGCCCTTTTTGGAAAT |
| ark1_F | GCAAATGGGAAGGAAAACAA |
| ark1_R | CTGAGAATCGTTTTGCACGA |
| taf6_F | GCTGCCCTATCGAGTTTACG |
| taf6_R | GATATTGGGAACGACGAGGA |
| gdh1_F | GCGTCTTCGCAATGAGTACA |
| gdh1_R | CTTCCTCACCGCTAACTTCG |
| mpg1_F | CCGATGATCCTTCACCAAGT |
| mpg1_R | CGAGGTAGGACGAGGTTCAA |
| **ChIP-qPCR:** | |
| q_Rad3_F | ACCGTATTTTAGCCGACGTG |
| q_Rad3_R | TAAAGGCGCTATTGGCACTC |
| q_Mac1_F | GCATTTCTAGGGCCATTTCC |
| q_Mac1_R | TACAAACATGACCGGTGCTC |
| q_ChrI_F | GGTTCATGATGGGTTCATCG |
| q_ChrI_R | ATATCCAACTGACGGCATCG |
| q_Hut1_F | TTCCGGAGAGACTGCTTTTG |
| q_Hut1_R | CTTTTCGTCTTCTCGGCATC |
| q_Pol5_F | TCCTCGTCTTTGACCTTTGC |
| q_Pol5_R | TCAAGATGCCTCTGACAACG |
| q_Pcn1_F | CGATGCCATCAAGGAGTTAG |
| q_Pcn1_R | GAAAGGCTTTGTCGTTGAGC |
| q_Ura4_F | TGCTGATATGCCTTCCAACC |
| q_Ura4_R | CATCATTGTTGGTCGTGGAG |
| q_Omh3_F | ACTGGCAGCATCGCTATTTC |
| q_Omh3_R | TGTACATCCGTAGGGCAATG |
| q_Mpg1_F | TGATGGTGTTCGTCTCCAAC |
| q_Mpg1_R | TCCAAACGACTCCAAGATCC |
| q_LTR1_F | AGGAAAACTCACCGCAGTTC |
| q_LTR1_R | CATCAATCCCATTGCCAAG |
| q_LTR2_F | AAGTCCAATCGGCAATAACG |
| q_LTR2_R | AGGAAAACTCACCGCAGTTC |
| q_LTR3_F | AGGAAAACTCACCGCAGTTC |

| | |
|---|---|
| q_LTR3_R | CTTCCTTATCAGGCCCAAAG |
| q_SV2NEO_F | TCTACGTGTTCCGCTTCCTT |
| q_SV2NEO_R | CCAGAAGTAGTGAGGAGGC |
| q_Mouse_Act_F | CCAAAGTAACAGGTCACTTACC |
| q_Mouse_Act_R | GTGTCTTGATAGTTCGCCATGG |
| q_Mouse_HoxC8_F | GCTCCCCAGACGAGTTCTGA |
| q_Mouse_HoxC8_R | GGCTGGGAGACGAGTTCTGA |
| q_MaSat_F | AAATACACACTTTAGGACG |
| q_MaSat_R | TCAAGTGGATGTTTCTCATT |
| q_Chr10_insert_0KB_F | CTCACAGGGTGTCTGCTCAA |
| q_Chr10_insert_0KB_R | AGGTTGTCAGGCTTGGTGAC |
| q_Chr10_insert_-1kb_F | GCGCATAAGCAGAGGTAAGG |
| q_Chr10_insert_-1kb_R | CTCTTCTCGGTCACCAGCTC |
| q_Chr10_insert_-2kb_F | CAGGTTCATCCCTGCTCAGT |
| q_Chr10_insert_-2kb_R | ACAAGATCAGGCGTGTCCTC |
| q_Chr10_insert_-10kb_F | CACACACCTGGAAACACCAG |
| q_Chr10_insert_-10kb_R | TATCTGTGTGCACGGATTGG |
| q_Chr10_insert_-100kb_F | CTAGCTAAGCGGCACAGACC |
| q_Chr10_insert_-100kb_R | AGTGTCAAAGGACGGTGGAG |
| q_Chr10_insert_-500kb_F | TTCTTCCATGGCTCTGGTTC |
| q_Chr10_insert_-500kb_R | TGGGTACCTTTCCACTCAGC |
| q_Chr10_insert_+1kb_F | ATTTGCTCAGTGTGCGTGAG |
| q_Chr10_insert_+1kb_R | TGACGTCAAGGCAGTGGTAA |
| q_Chr10_insert_+2kb_F | CAGAAGGGATAGGGCATTGA |
| q_Chr10_insert_+2kb_R | CTGGGTGATCATGGGAGAAG |
| q_Chr10_insert_+10kb_F | AGACTTGGGACAGGTGTTCG |
| q_Chr10_insert_+10kb_R | CAGCCATCTCAGGGACAACT |
| q_Chr10_insert_+100kb_F | CACGCCCAAATCTTGATAGG |
| q_Chr10_insert_+100kb_R | CGAGGCAAAACCTCTAGCAC |
| q_Chr10_insert_+500kb_F | GCTCAGGTGCCTTTTGTGTT |
| q_Chr10_insert_+500kb_R | GGTTCCAGGGTCTTCTGTCA |
| q_Human_Sat_F | AAACGGGGTTTCTTCCTTTC |
| q_Human_Sat_R | TGCCACAGCAAGAGTGTTTC |
| q_Human_GAPDH_F | TACTAGCGGTTTTACGGGCG |
| q_Human_GAPDH_R | TCGAACAGGAGGAGCAGAGAGCGA |
| **RT-qPCR:** | |
| qRT_ade5_F | CTGGATGGATGCACATTTTG |
| qRT_ade5_R | CGCGTTCGATAGCATGAATA |
| qRT_ima1_F | CTCAATTTGTGAAGCCACCA |
| qRT_ima1_R | GAAATCCCGTATCGCGTAAA |
| qRT_adh1_F | AGATTCCCGAAATGTGTCCA |
| qRT_adh1_R | AACATCGGCAAAGCAAAAAC |
| qRT_ade6_F | CATGGAAATTGCAGTGATGG |

| | |
|---|---|
| qRT_ade6_R | CGAGCAGGGGCATATACTAAA |
| qRT_ura4_F | TCGCAGACATTGGAAATACC |
| qRT_ura4_R | ATGGCAATTTGTGATATGAGC |
| qRT_SV2NEO_F | TCTACGTGTTCCGCTTCCTT |
| qRT_SV2NEO_R | CCAGAAGTAGTGAGGAGGC |
| qRT_Mouse_ActIntron_F | CCAAAGTAACAGGTCACTTACC |
| qRT_Mouse_ActIntron_R | GTGTCTTGATAGTTCGCCATGG |
| qRT_Mouse_HoxC8_F | GCTCCCCAGACGAGTTCTGA |
| qRT_Mouse_HoxC8_R | GGCTGGGAGACGAGTTCTGA |
| qRT_Mouse_CycA_F | TCGAGCTCTGAGCACTGGAG |
| qRT_Mouse_CycA_R | CATTATGGCGTGTAAAGTCACCA |
| qRT_Mouse_GAPDH_F | TTTGTCAAGCTCATTTCCTGGTATG |
| qRT_Mouse_GAPDH_R | AGTTGGGATAGGGCCTCTCTTG |
| qRT_Mouse_cMyc_F | TCAAGCAGACGAGCACAAGC |
| qRT_Mouse_cMyc_R | TACAGTCCCAAAGCCCCAGC |
| qRT_Mouse_Klf4_F | GGCGAGAAACCTTACCACTGT |
| qRT_Mouse_Klf4_R | TACTGAACTCTCTCTCCTGGCA |
| qRT_Mouse_Oct4_F | CCAACGAGAAGAGTATGAGGC |
| qRT_Mouse_Oct4_R | GTGCTTTTAATCCCTCCTCAG |
| qRT_Mouse_Sox2_F | TCTGTGGTCAAGTCCGAGGC |
| qRT_Mouse_Sox2_R | TTCTCCAGTTCGCAGTCCAG |
| qRT_Mouse_Eras_F | GCCCCTCATCAGACTGCTAC |
| qRT_Mouse_Eras_R | GCAGCTCAAGGAAGAGGTGT |
| qRT_Mouse_Tcl1_F | GCTTCCTCTCTGGGTGTTCA |
| qRT_Mouse_Tcl1_R | CCCACACATTCCCTTTCAAC |
| qRT_Mouse_Rex1_F | GGAAGAAATGCTGAAGGTGGAGAC |
| qRT_Mouse_Rex1_R | AGTCCCCATCCCCTTCAATAGC |
| qRT_Mouse_Gbx2_F | GCTGCTCGCTTTCTCTGC |
| qRT_Mouse_Gbx2_R | GCTGTAATCCACATCGCTCTC |
| qRT_Mouse_Nanog_F | CCTCCAGCAGATGCAAGAA |
| qRT_Mouse_Nanog_R | GCTTGCACTTCATCCTTTGG |
| **Bisulfite Sequencing:** | |
| Bisulf_mod21_F | AGAGGTAGAGTTTTTTTGTTATGTA |
| Bisulf_mod21_R | AAAATAATTTTACAAATTCATTTCC |
| Bisulf_arg3_F | AAGGATTTATTATGGTTATGAGAATT |
| Bisulf_arg3_R | TCAAAATTTATTTACACCAAACATC |
| Bisulf_mac1_F | TGTTAGTAATATTAGTTGTGGTAGGA |
| Bisulf_mac1_R | AATACTATAACTTCTCCAATACCATC |
| Bisulf_rpb10_F | ATTGTGTTTTATGTATGGTAGTTAT |
| Bisulf_rpb10_R | ACCTTTTATACCAATAAATTTTATC |
| Bisulf_cut2_F | TGATTGGTTTGTTTGGATTTATATG |
| Bisulf_cut2_R | AAAAAACACCAAACTTAAATTAACC |
| Bisulf_pol5_F | GTAGGGTTTTGTTTATTGATTTTTT |

| | |
|---|---|
| Bisulf_pol5_R | AACACCTTTTCATAATTTTTAAACC |
| Bisulf_ura4_F | TGAAATTAAAATATAGTAAAGTTTTG |
| Bisulf_ura4_R | TATAAACAAAAAATCAATTACCAAC |
| Bisulf_SV2NEO_F | AGGGTTTTTTAATTTTATTAGAGGG |
| Bisulf_SV2NEO_R | CCTAAACTTTTACAAAAAACTTCAC |
| Bisulf_gdh1_F | TTTTGTTATTATTTATGGGAAGGTT |
| Bisulf_gdh1_R | AAACTCCAAACTCTTTAAAATTCATC |
| Bisulf_omh3_F | TTGTTAAATGGTTATTGTTTATTAAA |
| Bisulf_omh3_R | AAAAAAAATCTTTTTCATCTACCTC |
| Strata_pSC-A_bGal_F1 | AATTAACCCTCACTAAAGGGAA |
| Strata_pSC-A_bGal_R1 | GTAATACGACTCACTATAGG |
| Strata_pSC-A_bGal_F2 | GGCTTTACACTTTATGCTCC |
| Strata_pSC-A_bGal_R2 | GCCAGCTGGCGAAAGGGGGA |
| Strata_pSC-A_bGal_F3 | TTTACGGTTCCTGGCCTTTT |
| Strata_pSC-A_bGal_R3 | GGCTGGGTGTGGCGGACCGC |
| **Cloning Primers:** | |
| Suv39_Chromo_F_SfiI | TAAGCAGGCCTCTGAGGCCGCAATGGCGGAAAATTTAAAA |
| JMJD2D_R_SfiI | TGCTTAGGCCTGACAGGCCTTAAACGGGCACAGGGGCCCA |
| EYFP_R_SfiI | TGCTTAGGCCTGACAGGCCTTACGTTTCTCGTTCAGCTTT |

# Chapter 3: Detailed characterisation of the *S. pombe* DNA insertion in the mouse F1.1 cell line

## 3.1 Introduction

The properties of the *S. pombe* DNA insert were last examined around twenty years ago using the approaches available at the time. Since then methods such as ChIP, next generation sequencing, 3C techniques and more have given us new ways to investigate the properties of chromatin. The application of such methods to the F1.1 cell line will provide insight into the nature and organisation of the *S. pombe* chromatin. However, proper analysis of the *S. pombe* chromatin by these techniques requires a more detailed knowledge of the DNA sequence that underlies it. Gaining a better understanding of the sequence composition and organisation of the *S. pombe* DNA is thus a key prerequisite to further study.

The sequence composition of the insert was previously examined to a certain extent by Southern analysis with specific probes for several *S. pombe* regions (McManus *et al.*). The presence or absence of certain chromosome regions allowed inferences to be made about the likely composition of the inserted *S. pombe* DNA. However, the rearranged nature of the insert DNA made it difficult to make robust conclusions. The presence of sequences from all three *S. pombe* chromosomes within the F1.1 genome suggests a complex composition of the insert which is difficult to resolve by indirect methods such as Southern analysis or Polymerase Chain Reaction (PCR). These indirect methods also leave open the question of the copy number of the sequences present within the insert, upon which calculations of the size of the insert must be based. Next generation Sequencing allows a more precise determination of the *S. pombe* DNA sequence and will facilitate experimental design for other approaches, including the choice of accurate primers for quantitative PCR.

A further open question is the precise location of the insertion of the *S. pombe* DNA within the mouse genome. Indeed, the potential influence of the surrounding mouse sequence and chromatin context on the *S. pombe* insert is a factor that must be investigated, as is the reverse question of the extent of the distinct nature of the chromatin at the insert and whether it extends beyond the limits of the foreign DNA

or is confined. For these reasons, determining the exact chromosomal location of the insertion is also crucial.

In this chapter I describe the sequence composition and location of the F1.1 insert. Complete sequencing and *de novo* assembly of the F1.1 genome was validated by PCR to reveal a much more complex composition insert than was previously supposed. Finally, FISH was used to confirm the exact site of *S. pombe* DNA insertion suggested by sequencing.

## 3.2 The F1.1 foreign DNA insert consists of 12 Mb of scrambled *S. pombe* DNA from across the genome

Microscopy based analysis previously estimated that the amount of *S. pombe* DNA inserted into the F1.1 chromosome was between 5 and 10 Mb. This estimate was derived by comparing the average length of the *S. pombe* insert with that of the whole chromosome across many images (McManus *et al.* 1994). In addition, Southern analysis with probes from regions across the *S. pombe* genome demonstrated that different sequences from all three *S. pombe* chromosomes were present and absent in F1.1. However, only 1 of the 5 probes for chromosome I and 2 of the 4 for chromosome II gave a positive signal in F1.1, while 7 out of 9 chromosome III regions were detected. However, centromeric sequences from all three chromosomes were detected. From this it was concluded that the F1.1 insert was a mix of DNA from several regions of the *S. pombe* genome, but likely consisted primarily of DNA from most of chromosome III and the distal arm of chromosome II, where almost all the identified regions mapped to.

In order to gain a clearer picture of the content and state of the DNA sequence within the *S. pombe* insert, Illumina whole genome sequencing was implemented to obtain the complete sequence of the F1.1 genome. Two independent F1.1 genomic DNA preparations were collected, from which paired-end tagged libraries were prepared and sent for sequencing. Reads were mapped either to the mouse genome, *S. pombe* genome, or with portions mapping to one of each. Both samples generated a high number of reads (279,061,718 and 277,167,713 respectively) with an overall alignment rate of 96.61% and 96.14%, respectively. Together the

samples covered 90.6% of the mouse genome, showing good agreement with each other.

**Figure 3.1** shows a graph of the F1.1 read counts from one sample mapped to the *S. pombe* reference genome. These results provide a more comprehensive picture of the *S. pombe* insert sequence, revealing a more complicated composition than previously supposed. Approximately 64% of the *S. pombe* genome is present to some degree in the F1.1 genome, as indicated by the read counts and represented schematically in the adjacent diagram. These regions are indeed scattered across the *S. pombe* genome, with all three chromosomes represented, however there is no over-representation of chromosome III as previously deduced. Moreover, no large regions of the genome (over 500kb) remain intact in the F1.1 cell line. Rather, the results indicate that the inserted DNA consists of a fusion of medium-sized DNA fragments, from a few tens to a few hundred kilobases, into one large stretch of DNA. While it is possible that some *S. pombe* DNA has been inserted separately into the F1.1 genome, previous and new FISH results (see section **3.3**) show only one visible region of *S. pombe* DNA, indicating that the majority of this foreign DNA has been inserted as a single entity into the mouse genome, and was likely recombined and fused prior to its insertion into the mouse genome.

The number of read counts mapped to each region of the *S. pombe* genome also provides information on their copy number within the F1.1 cell line. All regions of the *S. pombe* genome that were detected within the F1.1 genome in the sequencing data have read counts of either 200 or 400 approximately, indicating that they are present in either one or two copies, respectively. This then allows the total amount of *S. pombe* DNA present within the F1.1 genome to be calculated based on the length of the regions and their copy number. Approximately 5 Mb of DNA was found to be present in single copy, while 3.5 Mb was present in two copies. This gives a size of approximately 12 Mb for the *S. pombe* DNA inserted into the F1.1 genome, somewhat larger than the 5-10 Mb previously estimated.

While the sequencing results provide us with detailed information as to the presence or absence, as well as copy number, of certain regions of DNA, the sequencing depth was not of sufficient quality to *de novo* assemble the full sequence of the insert. Thus, although we can say to some degree of certainty that the sections of *S.*

A

| Sample | total reads (pair) | overall alignment rate | alignment rate | Perfect mapped | Intra-chromosome | Inter-chromosome |
|---|---|---|---|---|---|---|
| L7 | 279,061,718 | 96.61% | 93.54% | 257,029,604 | 2,076,390 | 1,928,337 |
| L8 | 277,167,713 | 96.14% | 92.59% | 251,608,023 | 3,230,448 | 1,791,114 |

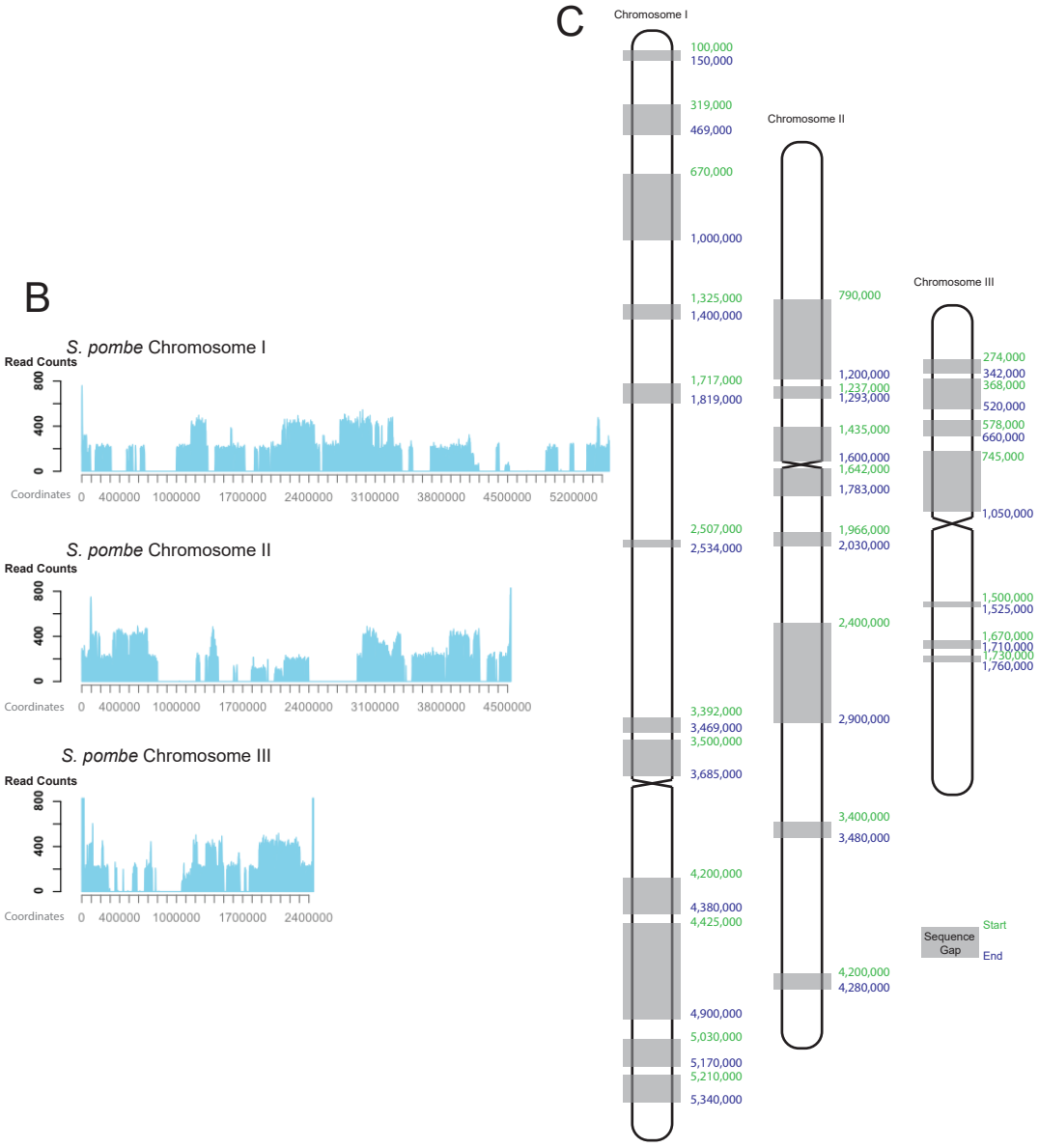


**Figure 3.1 - Sequencing results and map of the F1.1 *S. pombe* DNA content**

**A** - Details of the sequencing results of two paired-end libraries perpared from F1.1 genomic DNA
**B** - Sequencing reads from whole genome sequencing of F1.1 DNA mapped to the three chromosomes of the *S. pombe* genome.
**C** - Schematic representation of the DNA content of the F1.1 *S. pombe* insert, inferred from the sequence data. Regions devoid of any detectable reads are shown as greyed boxes (with approximate start and end coordinates), while those found to be present in F1.1 are left white.

*pombe* DNA we observe to be present in the insert will be largely intact, we are not able to ascertain the order in which these blocks are assembled within the F1.1 insert region.

In order to verify the DNA content of the *S. pombe* insert obtained by sequencing, PCR was performed using 24 primer pairs specific to *S. pombe* genes distributed over the three *S. pombe* chromosomes. Based on the sequencing results some were expected to be present in the F1.1 genome while others were expected to be absent in the F1.1 genome. PCR was performed on genomic DNA from F1.1, *S. pombe* DNA as a positive control and the parental mouse C127 cell line as a negative control (**Figure 3.2**). As expected, none of the genes absent in the F1.1 sequencing reads were amplified from F1.1 genomic DNA, while all of those present show a band at 1 kb. In contrast, all primer pairs amplify their target genes in *S. pombe* genomic DNA, while none of them do in the mouse genomic DNA.

The sequencing results, supported by PCR analysis, thus indicate that the foreign F1.1 insert is composed of approximately 12 MB of DNA from across the *S. pombe* genome, primarily consisting of a fusion of contiguous sections ranging from around 30 kb to a few hundred kb present at one or two copies, but in an undetermined order.

## 3.3 The foreign DNA is inserted at a specified locus in mouse chromosome 10

Previous FISH analysis indicates that the *S. pombe* DNA is primarily inserted as one large block into a single locus in the mouse genome, approximately two-thirds of the way down one of the longer mouse chromosomes. However, it is possible that small pieces of *S. pombe* DNA were also inserted at other locations in the mouse genome. Sequence analysis identified five regions of the mouse genome that may harbour *S. pombe* DNA. These were identified as hybrid junction reads in which one end mapped to the mouse genome and the other end mapped to the *S. pombe* genome (**Table 3.3**).

Of these five sites of potential *S. pombe* DNA insertion, four seemed unlikely to represent the site of primary insertion due to either the small size of the mouse

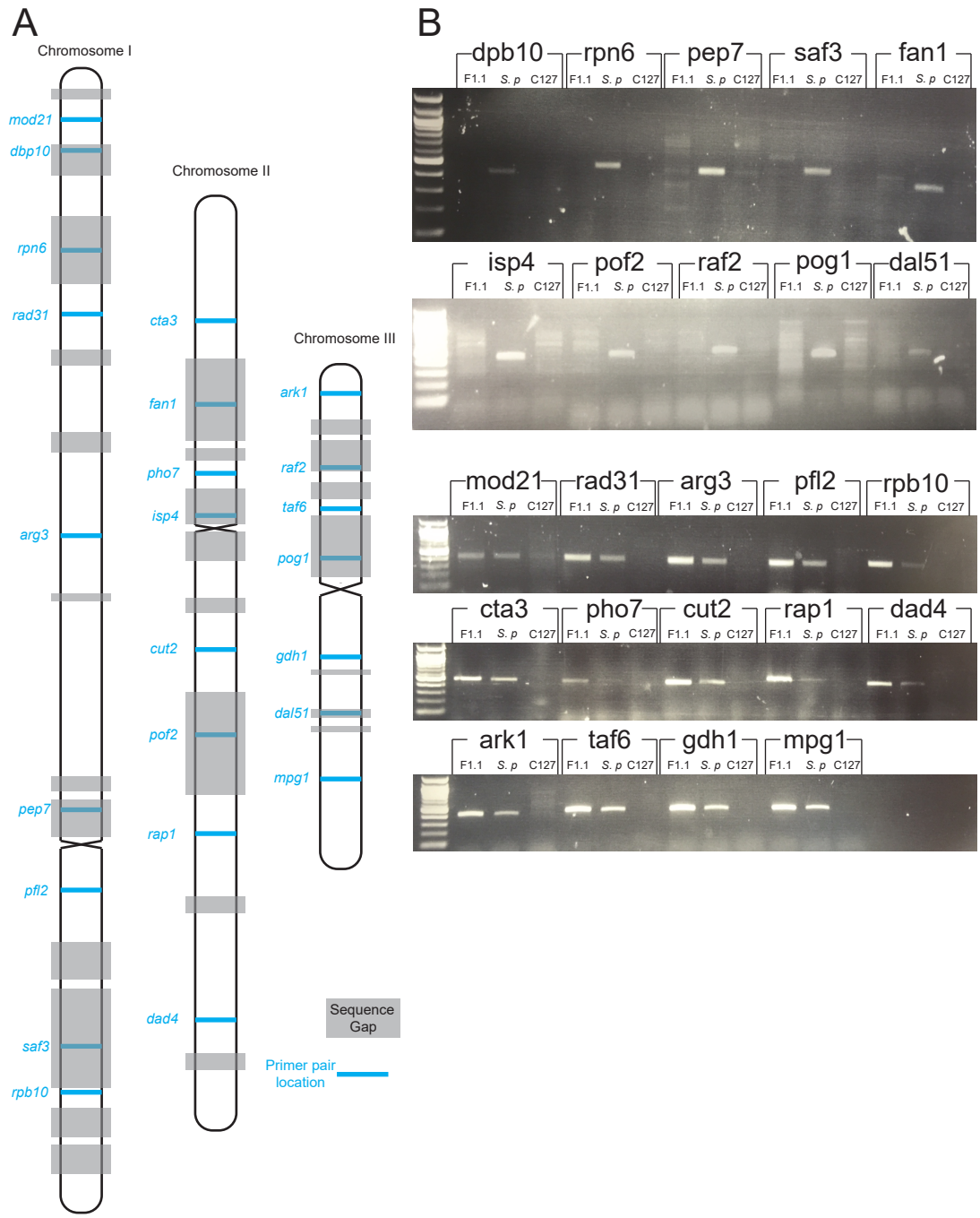


**Figure 3.2 - PCR confirms presence or absence of several *S. pombe* genes in F1.1 DNA**

**A** - Schematic representation of the DNA content of the F1.1 *S. pombe* insert, as determined in **Figure 3.1**. Regions found to be present in F1.1 are white while those found to be absent are greyed out. Location of primer pairs used in the PCR for **B** are indicated by blue bars. **B** - Confirmation of the sequence composition of the F1.1 *S. pombe* insert DNA by PCR. 24 PCR primers were designed to amplify 1 kb fragments from loci across the *S. pombe* genome found to be either absent (top two gels) or present (bottom three gels) in the F1.1 genome by whole genome sequencing. All primer pairs were used to amplify genome DNA from F1.1 (left), *S. pombe* (middle, positive control) and mouse C127 (right, negative control). Products were run on a 2% agarose gel stained with ethidium bromide.

| Mouse chromosome | Chromosomal region | S. pombe chromosome | Chromosomal region | Overlapping read count |
|---|---|---|---|---|
| 2 | 20,899,000-20,901,000 | III | 8,000-10,000 | 22 |
| | | III | 18,000-20,000 | 33 |
| 6 | 145,932,000-145,934,000 | III | 8,000-10,000 | 31 |
| | | III | 18,000-20,000 | 42 |
| | | III | 2,444,000-2,446,000 | 41 |
| 8 | 15,519,000-15,521,000 | III | 8,000-10,000 | 29 |
| | | III | 18,000-20,000 | 35 |
| | | III | 2,444,000-2,446,000 | 36 |
| 10 | 83,348,000-83,350,000 | II | 1,397,000-1,399,000 | 31 |
| 16 | 57,390,000-57,392,000 | III | 8,000-10,000 | 38 |
| | | III | 18,000-20,000 | 41 |
| | | III | 2,444,000-2,446,000 | 30 |

**Table 3.3 - Regions of apparent Mouse-*S. pombe* DNA junctions in the F1.1 cell line sequencing data**

chromosome on which they were located, the position of the hybrid junction site on the mouse chromosome not corresponding with the position observed by FISH, or both. Furthermore, these same four junction reads contained only *S. pombe* DNA that mapped to telomeric sequences in *S. pombe*. While it is possible that small pieces of *S. pombe* telomeric DNA were inserted at these locations in the mouse genome, the repetitive nature of telomeric sequences suggests that these reads may be an artefact of sequencing.

The remaining potential *S. pombe* DNA insertion site identified however fit all the criteria. This site corresponds to a region between coordinates 83,348,000 and 83,350,000 on mouse chromosome 10, which matched well with both the observed size of the chromosome that contains the insert and the position of the insert relative to it. Analysis of F1.1 cells is complicated by the fact that the parental C127 cell line does not have a normal mouse karyotype. In particular, most metaphase spreads of C127 and F1.1 show a chromosome number more consistent with a tetraploid genome and chromosome paints do indeed highlight more than two copies of a particular chromosome. In the case of chromosome 10, a chromosomal duplication / rearrangement also appears to have taken place, with the result that four full length chromosomes corresponding to chromosome 10 and two telomeric fragments are highlighted by a chromosome 10 paint in C127 (**Figure 3.4**). This is consistent with our sequencing data which show a duplication of the telomeric end of chromosome 10 (data not shown). However, the full-length copies of chromosome 10 represent the best candidates for the location of the *S. pombe* DNA insertion.

To test whether the *S. pombe* DNA is inserted in one of the copies of mouse chromosome 10, two-colour FISH was performed on metaphase spreads from the F1.1 cell line to highlight both mouse chromosome 10 and *S. pombe* DNA (**Figure 3.4**). The chromosome 10 paint (green) highlighted four copies of full-length chromosome 10. One of the copies contained a localised loss of signal in the green channel which coincided with the red signal indicating the *S. pombe* DNA insertion. Conversely, a chromosome 8 paint used to test one of the other candidate insertion sites highlighted several chromosomes with no obvious red *S. pombe* DNA signal in any. The presence of a red signal on another chromosome indicated that the FISH successfully detected the *S. pombe* DNA inserted elsewhere in the genome.

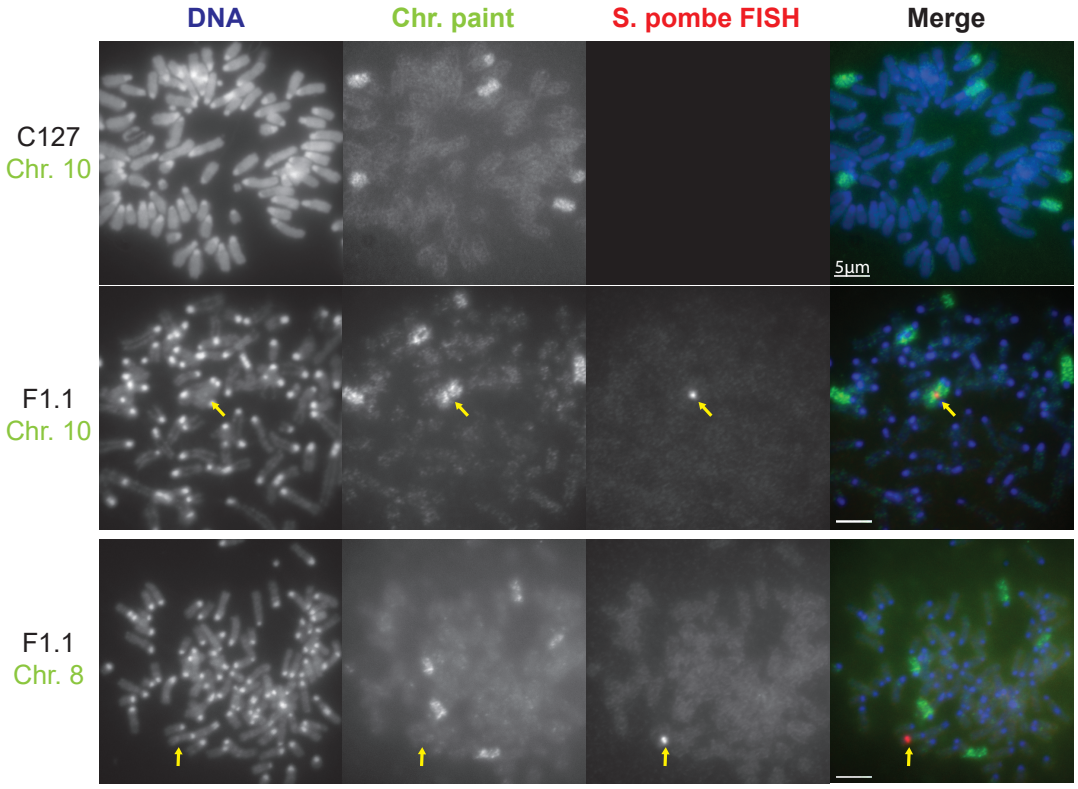


**Figure 3.4 The primary site of *S. pombe* DNA insertion in F1.1 is on mouse chromosome 10**

Two-colour FISH of F1.1 metaphase spreads using a Digoxygenin-labeled probe against total *S. pombe* DNA, detected with TRITC labeled anti-Digoxygenin antibodies (Red channel, middle-right), and a directly labeled FITC chromosome paint (Green channel, middle-left) for either mouse chromosome 10 (top 2) or 8 (bottom). DNA is labeled with DAPI (bottom 2) or PI (top) (Blue channel, far left). Insert location is indicated by yellow arrows. Note that for the top images of C127 *S. pombe* FISH was not performed and the red channel is thus blank. Scale bar indicates 5 µm.

These FISH analyses, coupled with whole genome sequencing, clearly demonstrate that the majority of the *S. pombe* DNA is inserted at a single locus in the F1.1 cell line, corresponding to a region between coordinates 83,348,000 and 83,350,000 on one of the copies of mouse chromosome 10. A genome browser view of this region within a 200kb window is shown in **Figure 3.5**. The insertion site corresponds to an intergenic locus between two quite ubiquitously expressed mouse genes: the ion-channel protein Slc41a2 and the uncharacterised D10Wsu102e gene. The insertion does not disrupt either of these open reading frames, although it is within the transcription site of a putative intergenic lncRNA. ChIP-seq data from the ENCODE project shows the expected pattern of histone modifications over the two genes: H3K36me3 levels are high over the whole genome bodies while H3K4me3 and H3 acetylation show peaks at the transcription start sites of both genes. However, no distinctive pattern is seen over the *S. pombe* insert region, with both activating and repressive histone modification levels at or near background across the 2kb insert window.

## 3.4 Discussion

The sequence composition of the *S. pombe* DNA insert in the F1.1 cell lines was determined by whole genome sequencing of F1.1 DNA (**Figure 3.1**). The insert was found to be composed of a seemingly random assortment of regions from across the *S. pombe* genome, totalling approximately 12 Mb of sequence. These regions were found to be represented at no more than two copies, confirming that the distinct chromatin structure is assembled on essentially unique sequence. The *S. pombe* DNA in F1.1 thus provides a valuable tool for the study of distinct chromosomal structures, unhindered by the obstacles inherent to the repetitive sequences that make up the distinct structures observed at endogenous centromeres.

However, the *S. pombe* DNA insert remains complex in other respects. The mixed composition of the inserted sequence confirms previous suspicions that the *S. pombe* DNA was rearranged prior to its insertion in the mouse genome. To date the sequencing performed has not been of sufficient depth or length to allow the *de novo* assembly of an ordered map of the *S. pombe* DNA content of F1.1. This lack of a clear picture of the *S. pombe* DNA will complicate the analysis by certain methods such as Hi-C, making it difficult to assess datasets spanning large regions.

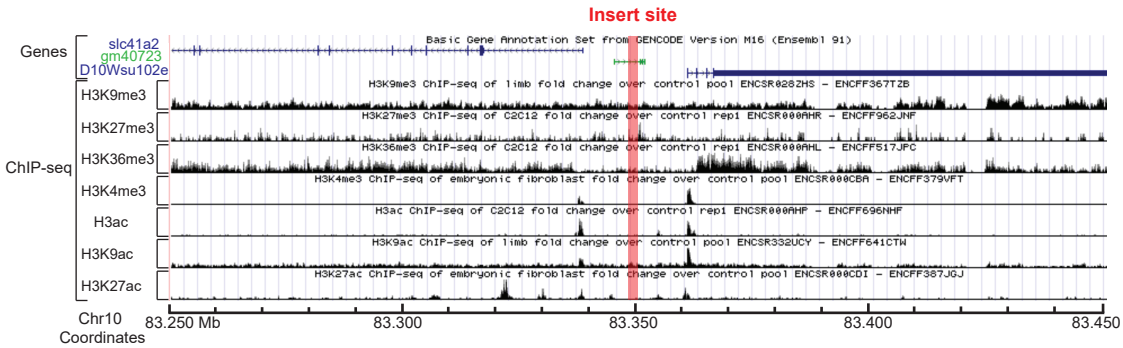


**Figure 3.5 - The *S. pombe* DNA is inserted in an intergenic region with no distinguishing chromatin marks**

UCSC genome browser view of the site of *S. pombe* DNA insertion on mouse chromosome 10 in the F1.1 cell line, with 100kb of flanking regions on either side. The site of *S. pombe* DNA insertion is located in the exact middle and indicated by a red bar. Gene open reading frames are indicated in blue and lncRNAs in green. ChIP-seq data from the ENCODE project is plotted in black as fold change over control for several histone modifications. Wherever possible data from MEFs or the C2C12 cell line was used, otherwise data from whole embryonic limbs was used.

However, the sequencing results do suggest the existence of a few sizeable segments of contiguous DNA with respect to the original *S. pombe* genome. For instance, the reads mapping to Chromosome II 3,500,000-3,800,000 and Chromosome III 1,900,000-2,300,000 show no break in continuity nor change in copy number, strongly suggesting contiguous regions of 300 and 400 kb, respectively. Analysis of datasets over these sequences can be done with high levels of confidence that they represent single large regions. Knowledge of the exact content of the insert has also facilitated the design of other experiments, for instance the design of qPCR primers for ChIP of sequences known to be present within the insert.

Based on the sequencing data a likely candidate for the insertion site was identified and confirmed by FISH using mouse chromosome paints (**Figure 3.4**). The primary site of insertion was shown to be on mouse chromosome 10 between coordinates 83,348,000-83,350,000. Four other potential insertion sites were flagged by the whole genome sequencing. However, in all of these cases the *S. pombe* DNA sequence identified as forming a junction with mouse DNA was telomeric, and thus repetitive. This raises the possibility that these potential mouse- *S. pombe* junction reads arose from some kind of artefact of the sequencing process. FISH analysis of F1.1 highlighting *S. pombe* DNA has always shown only one primary insertion to be evident. In addition, two-colour FISH showed no detectable *S. pombe* DNA signal on mouse chromosome 8 (the location of one of the four possible insertion sites). The amount of DNA inserted at these other locations is therefore either undetectably small or non-existent.

The results thus clearly demonstrate that the primary site of *S. pombe* DNA insertion is at a single locus on mouse chromosome 10. The insertion site does not disrupt any open reading frame nor does anything stand out about its chromatin profile from publicly available ChIP-seq data (**Figure 3.5**). This information should be kept in mind as we proceed to investigate the chromatin marks of the *S. pombe* chromatin in **Chapter 5**. It seems unlikely that the surrounding chromatin is influencing the histone modification levels of the *S. pombe* insert, given that it does not have a strong chromatin profile, either repressive or permissive. On the other hand, it will be important to determine whether the reverse is true and the *S. pombe* chromatin is affecting the histone modifications of its flanking regions in some way.

Together, these results thus provide a useful framework for further analysis of the structure and properties of the *S. pombe* chromatin inserted in the F1.1 cell line.

# Chapter 4: The foreign chromatin insert adopts a constricted appearance and a different chromatin organisation in mitosis

## 4.1 Introduction

The observation of an unusual structure associated with *S. pombe* DNA inserted in a mouse chromosome in the F1.1 cell line has so far relied primarily on microscopy. While the structures seen by electron microscopy are compelling, it is difficult to draw conclusions as to the internal organisation of chromatin within the chromosome from these exterior views. Furthermore, light microscopy has thus far shown the structure to be unusual only through visualisation of the DNA with stains such as propidium iodide. Detailed examination of the distinctive attributes of the unusual chromatin insert beyond its manifestation in the appearance of DNA is therefore necessary. These analyses must also consider current models of mitotic chromosome folding and how such models might explain the structure exhibited by the foreign insert. Given our increased understanding of the role of chromatin organising proteins such as condensins in mitotic chromosome architecture, an essential step towards new insights will be to gain a better understanding of the relationship between the chromatin fibres and these proteins. Another important aspect will be to understand the internal organisation of the chromatin in a manner that is beyond the wholly visual approach of microscopy. For these purposes, modern chromosome conformation capture techniques, 3C and its derivates such as Hi-C, provide an excellent method to probe such internal organisation, as they have already been used to provide insight into the organisation of mitotic chromosomes (Naumova *et al.* 2009, Gibcus *et al.* 2018).

The F1.1 cell line was recovered from a few cells that remained viable after storage in liquid nitrogen for twenty years. Therefore, a prerequisite for any further analyses was to determine if the *S. pombe* insert still presented the same unusual structure in the recovered cells as it did previously. Furthermore, if conclusions about mitotic chromosome structure in general are to be drawn from this particular example it is also important to determine if the structure adopted by *S. pombe* DNA in the F1.1 cell line is an exception or if it is a general property of foreign chromatin integrated into chromosomes of any cell type. While other examples of cells displaying unusual structures at foreign chromatin were originally presented in McManus *et al.* they

were all products of the same cell line. Testing the reproducibility of the formation of these structures and extending them beyond a single cell type will confirm that the observed phenomenon is a product of more general cellular processes. The mechanism that mediates such structural differences can then be investigated.

With these goals in mind, the experiments detailed in this chapter set out to:

1. Determine if the *S. pombe* insert of the F1.1 cell line still exhibits an unusual mitotic chromosome structure, and extend observations of it beyond that of DNA alone.
2. Insert large blocks of *S. pombe* chromatin into other mammalian cell lines and determine if they exhibit an unusual chromatin structure.
3. Investigate the properties of chromatin organising proteins at the *S. pombe* insert and how they are interacting with the chromatin fibres to determine if and how they are influencing the structure of the chromatin there.

## 4.2 DNA and histone levels are depleted at the region of foreign chromatin insertion

It was previously found that the *S. pombe* foreign DNA insertion in F1.1 adopted a distinct structure on the metaphase chromosome spreads (McManus *et al.* 1994). Following propidium iodide (PI) staining a narrowing of the chromosome width and darkening of the DNA stain was observed over the region of *S. pombe* DNA insertion. This distinctive structure remains evident in metaphase spreads of the recovered F1.1 cell line as shown in the gallery of selected images (**Figure 4.1 A**). The average PI signal intensity was measured at various points along the length of the insert-bearing chromosome 10 from the telocentric centromere to the distal telomere. The averaged intensities show a significant dip in signal strength in a region corresponding to 65-80% of the total chromosome length. This position coincides exactly with the region where foreign DNA is inserted as indicated by the peak in the *S. pombe* DNA FISH signal (**Figure 4.1 B**).

To rule out the possibility that the FISH procedure or signal interferes with the DNA stain signal, the same experiment was performed using a mouse Chromosome 10 fluorescent paint. Since the *S. pombe* DNA should be the only region of the

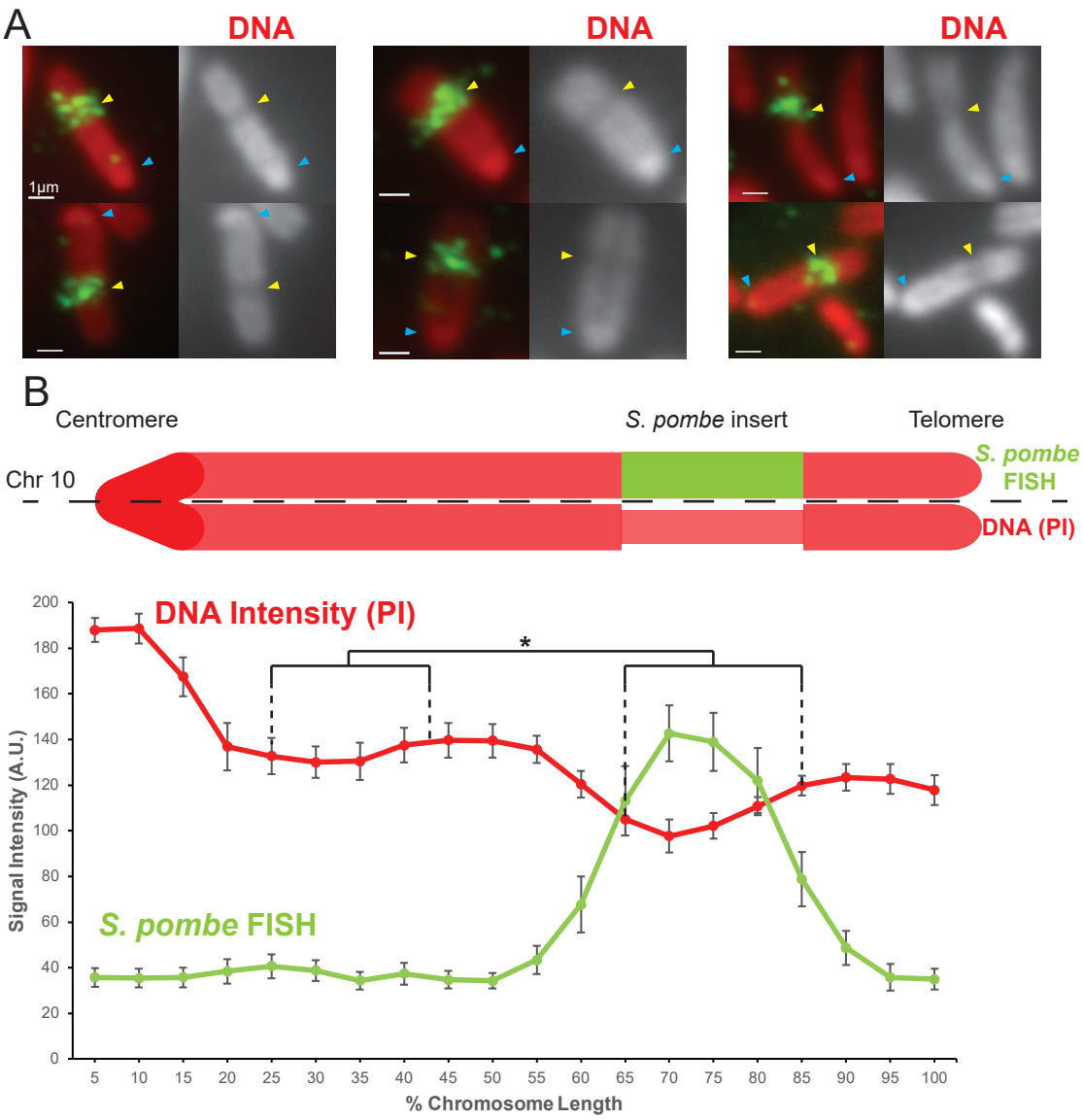


**Figure 4.1 - A distinct metaphase chromosome structure is detectable over the *S. pombe* DNA in F1.1**

**A** - Metaphase spreads of F1.1 chromosomes showing the distinct structure of the foreign chromatin insert. DNA is stained with propidium iodide (PI) and *S. pombe* DNA is highlighted by FISH using biotinylated total *S. pombe* DNA bound to avidin-FITC and indicated by yellow arrows. Centromeres appear as regions of bright DNA staining and are indicated by blue arrows. Scale bars indicate 1µm.
**B** - Average chromosome profile of the F1.1 insert bearing chromosome across several images (n=16). Signal intensities of PI DNA stain (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position, from the centromere (0-5%) to the telomere (95-100%), before plotting. Error bars represent +/- standard error from the mean (SEM). Average DNA stain intensity was compared between the regions of 25-45% (endogenous mouse DNA) and 65-85% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01).

chromosome not highlighted by the chromosome paint this experiment should determine if the unusual structure of the *S. pombe* chromatin is an artefact of FISH. The chromosome 10 paint also allows direct comparison of the insert-bearing chromosome 10 and its homologous counterpart that lacks the foreign DNA insert. The resulting microscopy images clearly show the same distinct appearance of an insert on one of the copies of chromosome 10 (**Figure 4.2**). Quantification confirmed that a region of decreased PI intensity corresponded with a decrease in chromosome paint intensity at a position around 65-85% of the total chromosome length (**Figure 4.3**). In contrast the other copies of chromosome 10 in F1.1 and all chromosome 10 homologs of the parental C127 cell line show no variation in the intensity of either DNA or chromosome paint intensity along their entire length, except for an expected increase in DNA signal at the centromere regions.

PI binds to DNA in a sequence-independent manner (Martin *et al.* 2005). The intensity of the PI stain measured by microscopy can therefore be taken as an indicator of the amount of DNA present over a given area. The clear decrease in PI intensity across the *S. pombe* DNA insert suggests that the amount of DNA per unit length is lower across the *S. pombe* insert than in the surrounding endogenous chromatin on the metaphase chromosome. This is consistent with the observed narrowing of the chromosome over the *S. pombe* DNA. Taken together, these results suggest not that the chromatin is more compact over the *S. pombe* DNA (which would result in an increase in PI stain) but that its organisation is altered such that the chromosome width is decreased without altering the compaction level of the chromatin. In this manner the amount of chromatin decreases concomitant with a decrease in chromosome width.

To confirm this, the levels of other chromatin components at the *S. pombe* insert were assessed. First, the levels of histone H4 were examined by immunolocalisation (**Figure 4.4**). Just as with the DNA stain, the H4 signal showed a significant decrease in intensity coinciding exactly with the site of *S. pombe* DNA FISH. This decrease in histone levels supported the idea that overall chromatin is decreased at the *S. pombe* DNA insert. However, given that the *S. pombe* chromatin may adopt an unusual structure, it remained possible that the decrease in apparent H4 signal was due to the altered chromatin structure preventing access of the H4 antibody to its target histone. To rule out the possibility of the *S. pombe* chromatin structure

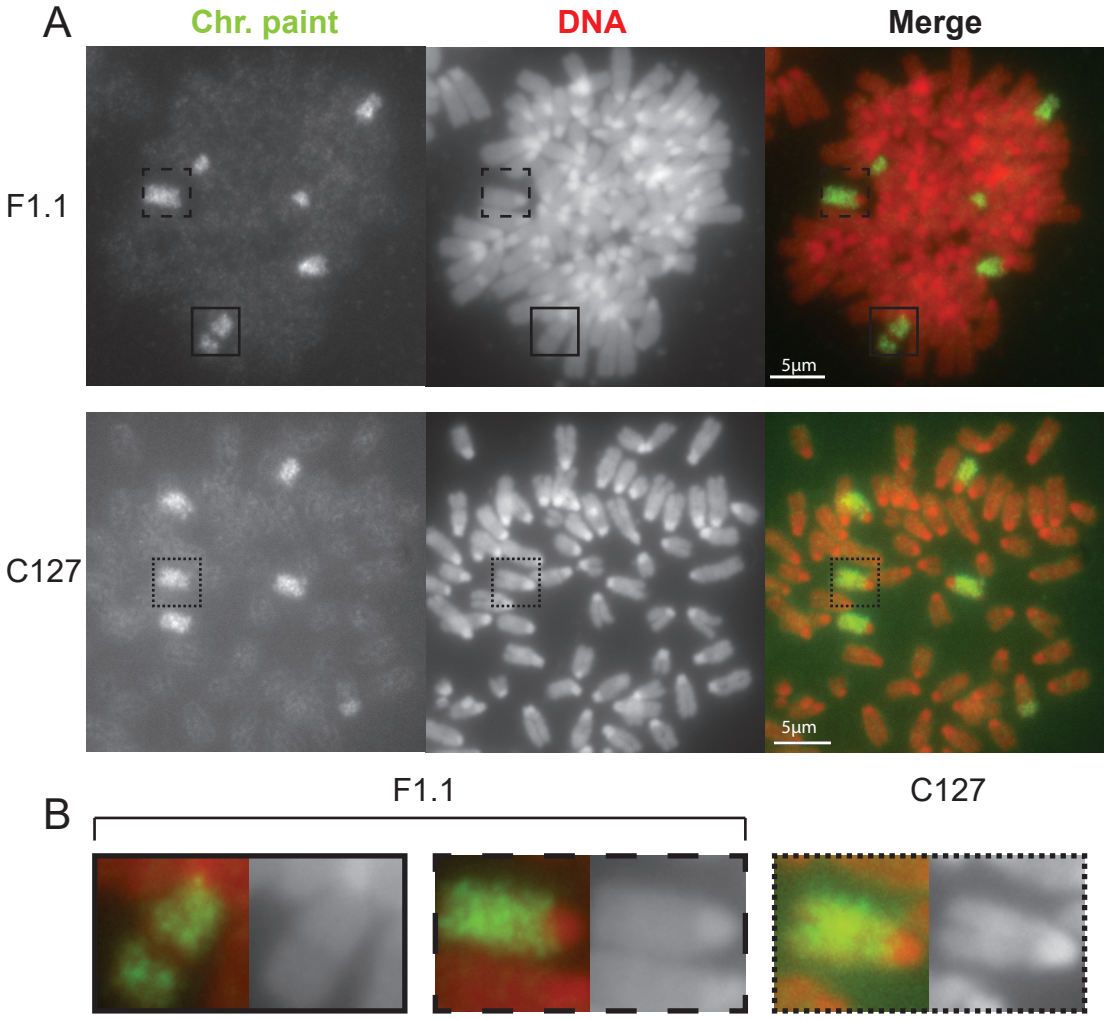


**Figure 4.2 - The distinct chromatin structure of the *S. pombe* DNA insert on chromosome 10 is not a product of the FISH procedure**

**A** - Metaphase spreads of F1.1 (top) and C127 (bottom) probed with a mouse chromosome 10 paint (green channel) and stained with PI (red channel). In F1.1, the insert-bearing chromosome 10 is recognisable by a distinct region not highlighted by FISH, corresponding to the *S. pombe* DNA. Scale bars represent 5µm.
**B** - Enlarged images of the regions highlighted in **A** showing the insert-bearing F1.1 chromosome 10 (left) as well as non insert-bearing copies of chromosome 10 from F1.1 (middle) and the parental C127 cell line (right).

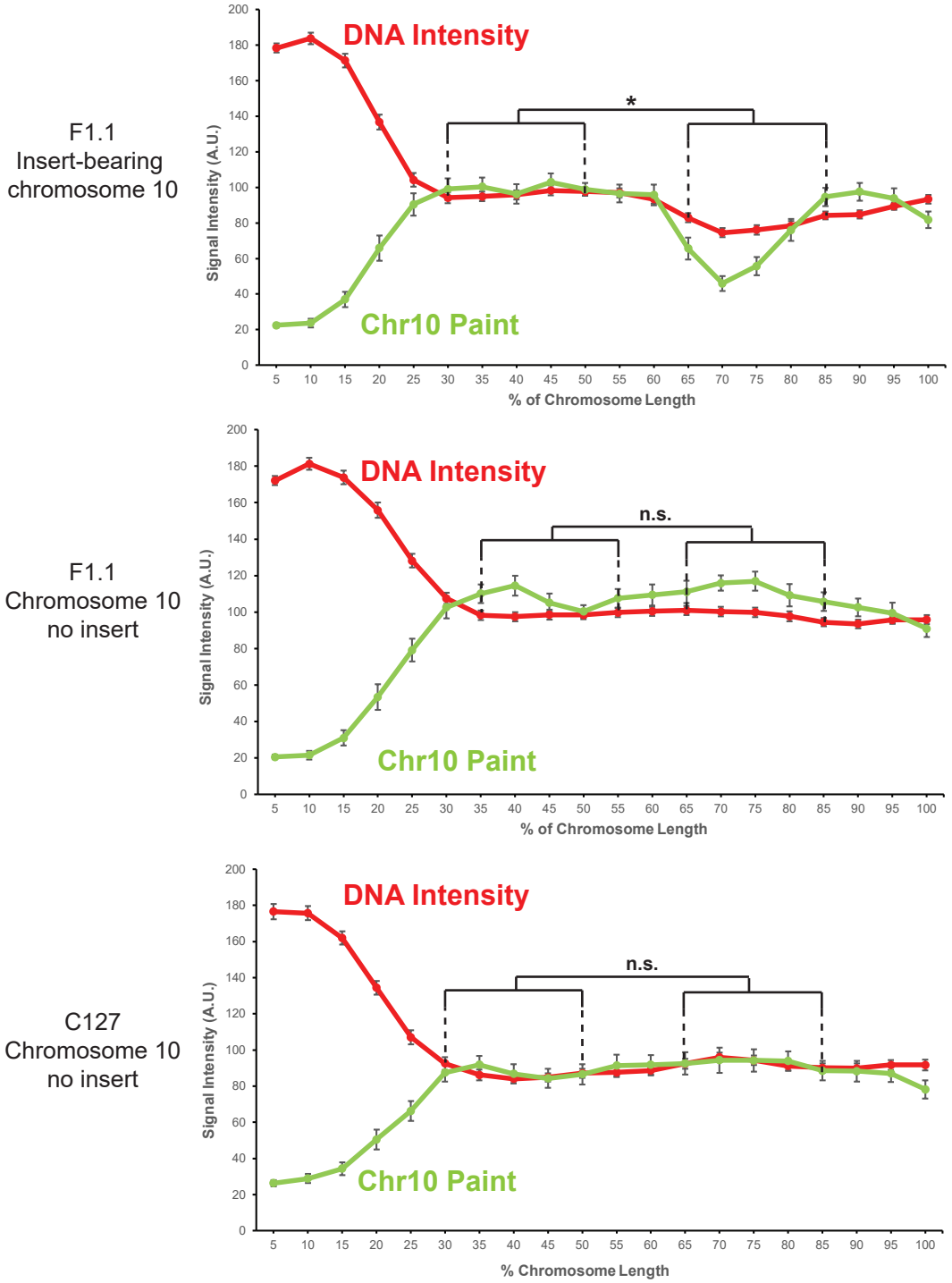


**Figure 4.3 - Quantification of the chromatin structure of insert-bearing and non insert-bearing copies of chromosome 10 in F1.1 and C127**

Average chromosome profile of the insert-bearing and homologous non insert-bearing copies of chromosome 10 in F1.1 and C127 metaphase spreads across several images (n=38, 38 and 39, respectively). Signal intensities of PI DNA stain (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- standard error from the mean (SEM). Average DNA stain intensity was compared between the regions of DNA at 30-50% or 35-55% and 65-85% (where the S. pombe DNA was detected on the insert-bearing chromosome) by the KS test (*p<0.001, n.s. = not significant).

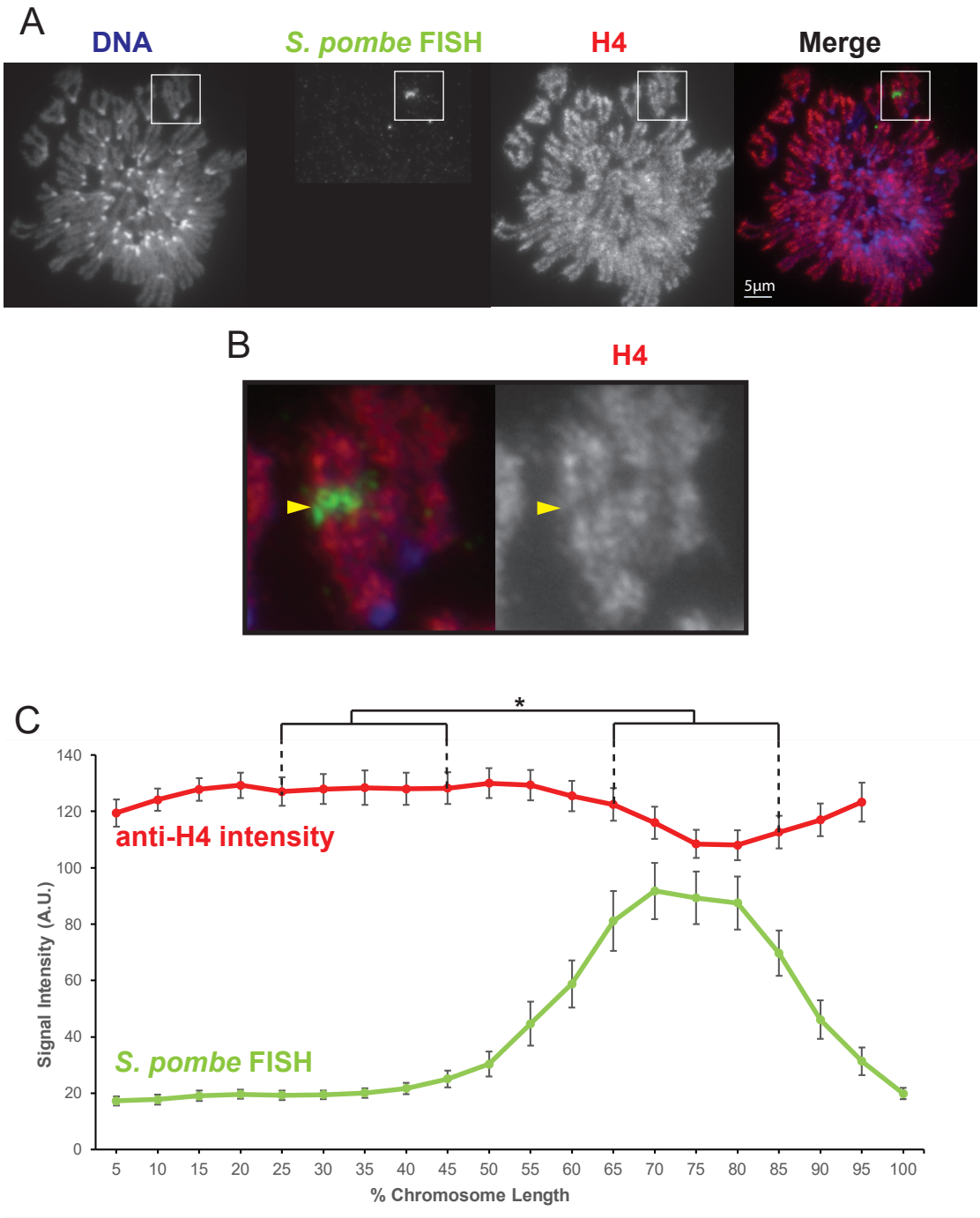


**Figure 4.4 - Histone H4 levels are reduced at the *S. pombe* DNA insert in F1.1**

**A & B** - Immunolocalisation of F1.1 metaphase spreads for histone H4 (red channel) with FISH for *S. pombe* DNA (green channel). DNA is stained with DAPI (blue channel). Scale bar indicates 5 µm. The insert-bearing chromosome is highlighted in **A** and shown enlarged in **B**, with yellow arrows indicating the site of *S. pombe* DNA.
**C** - Average chromosome profile of the F1.1 insert bearing chromosome across several images (n=28). Signal intensities of H4 IF (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average H4 intensity was compared between the regions of 25-45% (endogenous mouse DNA) and 65-85% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01).

interfering with the immunolocalisation procedure, a transgenic H2B histone directly tagged with the fluorescent GFP marker was transiently expressed in F1.1 cells and its levels on the mitotic chromosome assessed by microscopy (**Figure 4.5**). Once again, H2B-GFP signal intensity decreased at the site of *S. pombe* DNA insertion compared with the surrounding mouse chromatin, in a similar manner to H4 and DNA signals.

Analysis of the DNA stain intensity thus confirms previous observations of a distinct chromatin organisation over the *S. pombe* DNA insert in F1.1 metaphase chromosomes and is supported by the finding that histone levels also decrease across the *S. pombe* insert. Together these observations indicate that a decrease in the amount of chromatin per unit length occurs over the region of *S. pombe* chromatin compared with endogenous mouse mitotic chromatin.

### 4.3 *S. pombe* DNA inserted into mammalian chromosomes following fusion with other cell lines also exhibit a similar distinct chromatin structure

The distinct appearance of the *S. pombe* chromatin on chromosome 10 in F1.1 may be a rare event that is not reproducible. In order to determine if a similar structure to that observed in F1.1 results following *S. pombe* insertion into a different cell line, the same SV2NEO containing strain of *S. pombe* was fused with mouse NIH-3T3 cells. Just as for the generation of F1.1, spheroplasts were prepared from SV2NEO-pombe and fused with PEG with the recipient NIH-3T3 cells. After selection for G418 resistance two cell lines emerged. One of them had no visible insertion of *S. pombe* DNA as assessed by FISH. It is likely that this cell line retained only a small region of *S. pombe* DNA containing the SV2NEO gene. However, the other cell line (named NP-A4) exhibited a single large *S. pombe* DNA insertion on the arm of one of its chromosomes (**Figure 4.6**). Intensity measurements revealed that the PI signal decreased at a position around 75-85% along the length of the chromosome corresponding to the site of *S. pombe* DNA insertion as visualised by FISH. Thus, just as in F1.1, the *S. pombe* chromatin insert exhibits a constricted appearance.

*S. pombe* spheroplasts were then fused again by the same method to cells from the human HeLa cell line. Again, after fusion and selection two cell lines emerged as being G418 resistant, and were named HeP-2 and HeP-3. This time both cell lines

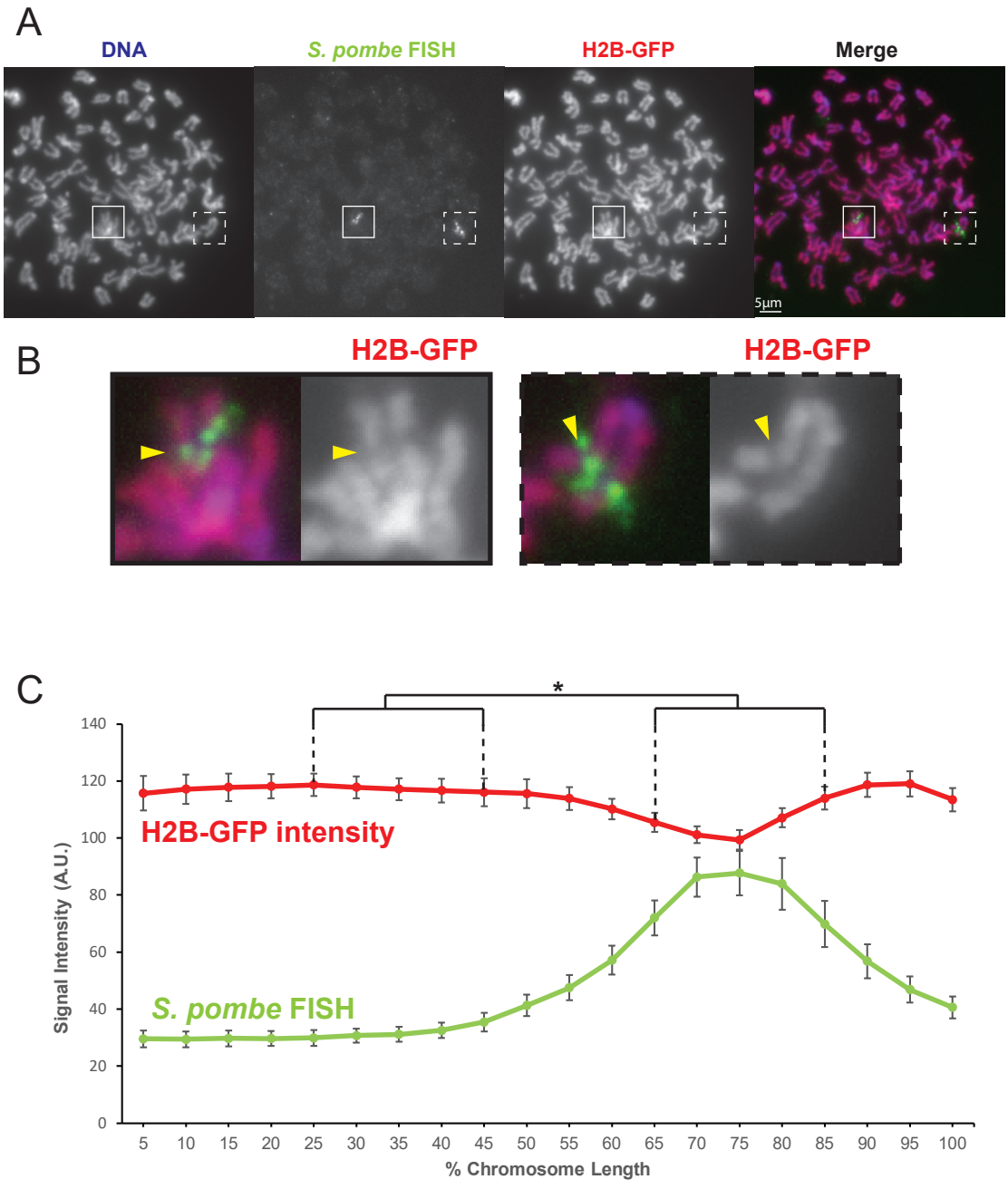


**Figure 4.5 - Direct measurement of tagged histone H2B levels shows a reduction at the *S. pombe* DNA insert in F1.1**

**A & B** - Metaphase spreads of F1.1 with fluorescently tagged histone H2B (red channel) and FISH for *S. pombe* DNA (green channel). DNA is stained with DAPI (blue channel). Scale bar indicates 5 µm. Insert-bearing chromosomes are highlighted in **A** and shown enlarged in **B**, with yellow arrows indicating the site of *S. pombe* DNA.
**C** - Average chromosome profile of the F1.1 insert bearing chromosome across several images (n=27). Signal intensities of H2B-GFP (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average H2B-GFP intensity was compared between the regions of 25-45% (endogenous mouse DNA) and 65-85% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01).

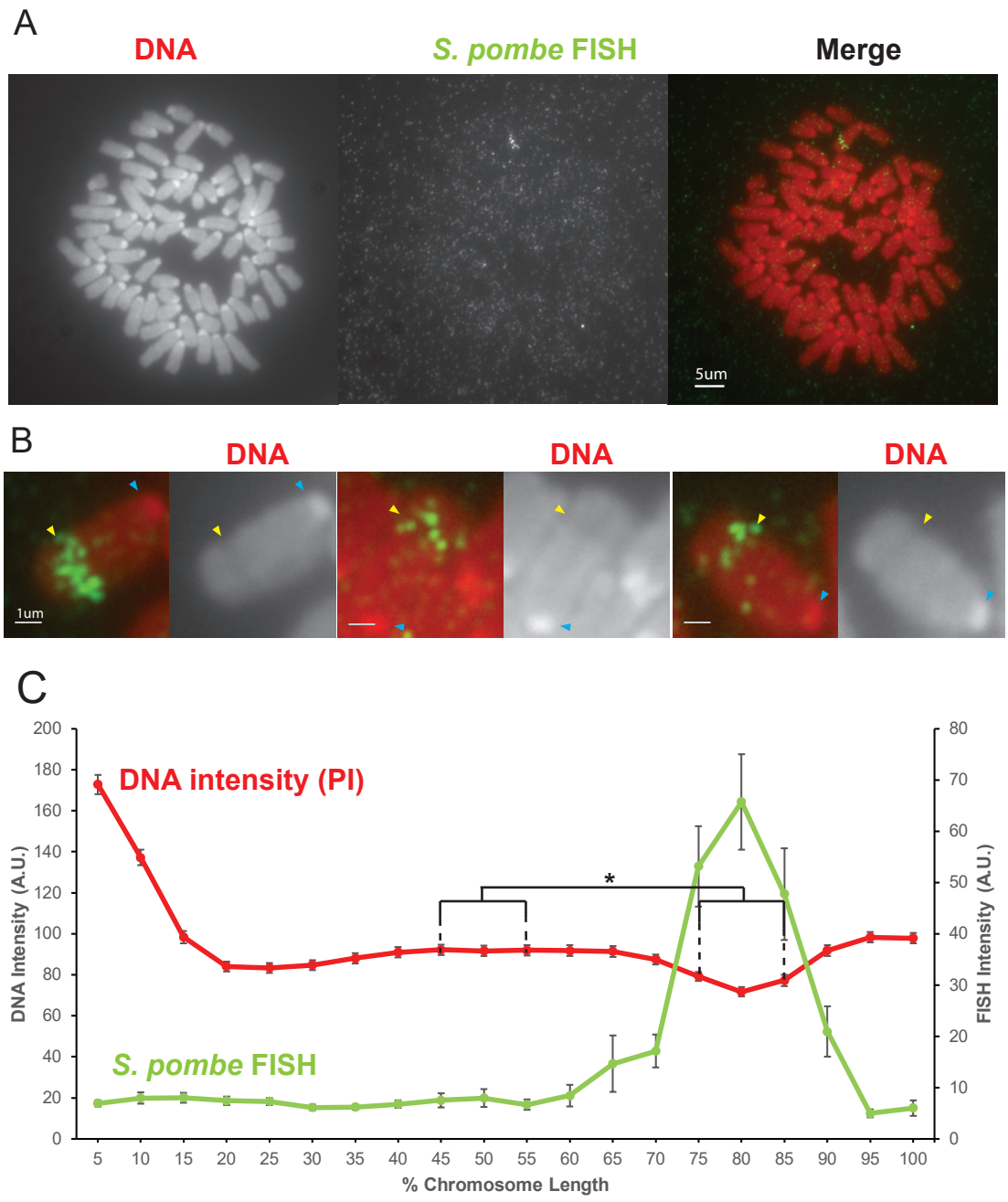


**Figure 4.6 - A distinct chromatin structure forms on *S. pombe* DNA inserted in a mouse NIH-3T3 cell chromosome**

**A & B** - Metaphase spreads of NP-A4 chromosomes showing the insertion of a large amount of *S. pombe* DNA with a distinct structure. DNA is stained with propidium iodide and S. pombe DNA is highlighted by FISH and indicated by yellow arrows. Centromeres appear as regions of bright DNA staining and are indicated by blue arrows. Scale bars indicate 5 μm (in **A**) or 1 μm (in **B**).
**C** - Average chromosome profile of the NP-A4 insert bearing chromosome across several images (n=37). Signal intensities of DNA stain (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average DNA stain intensity was compared between the regions of 45-55% (endogenous mouse DNA) and 75-85% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01).

exhibited regions of visible *S. pombe* DNA by FISH. In HeP-2 all the *S. pombe* DNA appeared to be present on a minichromosome with little or no surrounding human DNA, making analysis of its structure difficult (**Figure 4.7**). However, in HeP-3 a large insert of *S. pombe* DNA was visible within the arm of a human chromosome (**Figure 4.8**). Once again, this foreign *S. pombe* chromatin displayed the same distinct structure, as seen by a decrease in PI signal intensity in the region marked by *S. pombe* DNA FISH. Furthermore, unlike in acrocentric mice, human chromosomes have a distinct centromeric constriction. In HeP-3 the centromere and *S. pombe* insert had very similar appearances along the length of the chromosome.

These two additional cell lines show that large blocks of *S. pombe* DNA inserted into host chromosomes of other mammalian cells will adopt a similar distinct structure to that observed in the F1.1 cell line, derived from mouse C127 cells. This demonstrates that the unusual structure is not a rare event, limited to F1.1, but is a general property of *S. pombe* chromatin inserted into mouse or human chromosomes.

## 4.4 Condensin levels remain constant throughout the chromatin region containing *S. pombe* DNA

The analysis presented in **Section 4.2** showed that both DNA and histone levels decrease across the *S. pombe* chromatin on the insert-bearing chromosome 10 of F1.1. This suggests an altered organisation of the chromatin over this region of the chromosome. As key proteins in the process of chromatin organisation, the levels of the condensin complexes were assessed next.

The levels of the SMC2 protein, a primary component of both condensin complexes I and II, was assessed along the length of the insert-bearing chromosome 10 in F1.1 by immunolocalisation (**Figure 4.9**). As expected, condensin localised axially along the whole length of all mitotic chromosomes. This pattern was maintained over the whole insert bearing chromosome 10, showing no perceptible variation in SMC2 intensity within the *S. pombe* chromatin insert. Indeed, the intensity of the SMC2 signal remains remarkably stable across the entire length of the chromosome, with the exception of the centromeric region, and no significant difference was observed between the region of *S. pombe* DNA and the rest of the chromosome arm.

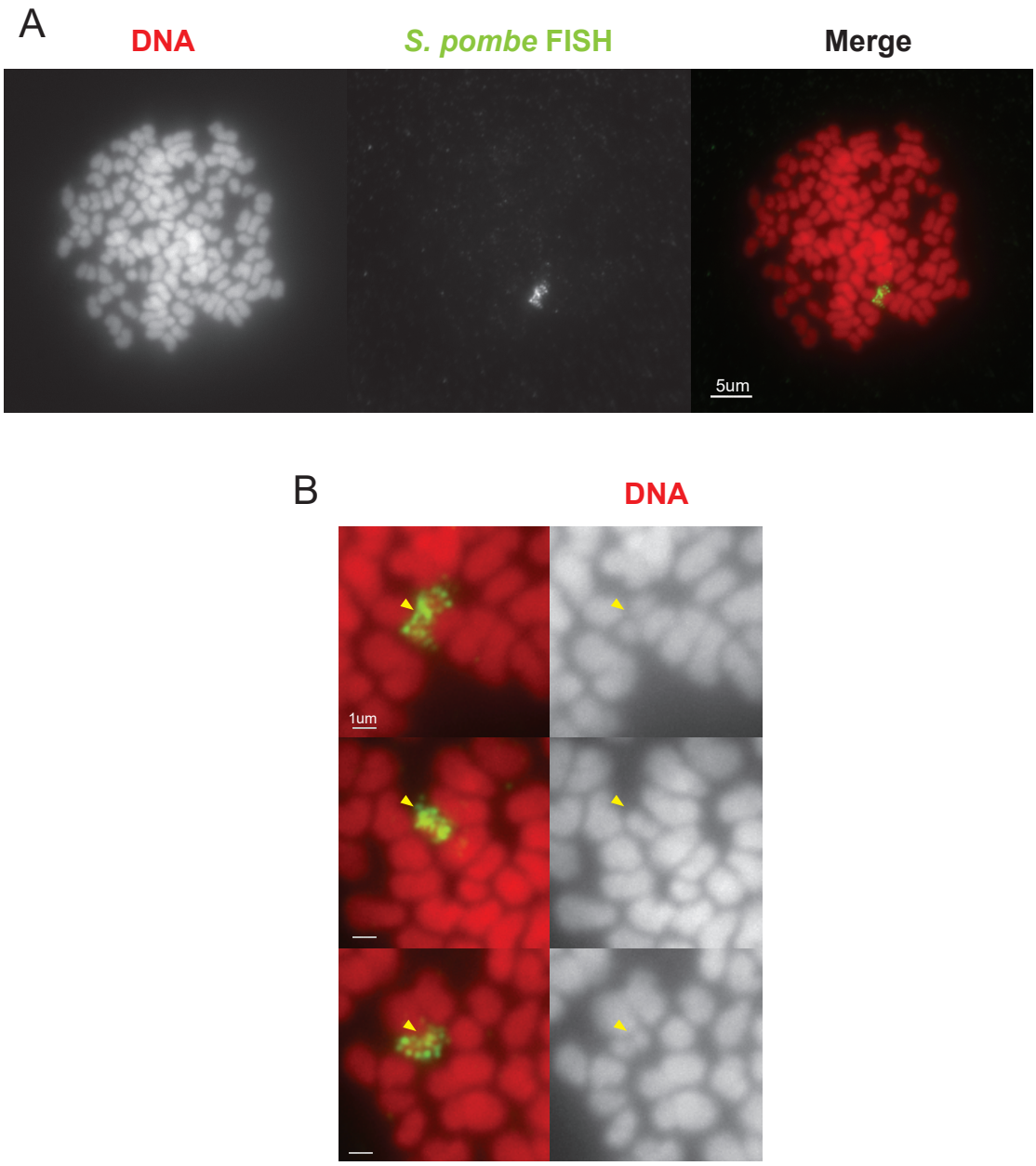


**Figure 4.7 - *S. pombe* DNA appears as a minichromosome in the human HeP-2 cell line**

**A & B** - Metaphase spreads of HeP-2 chromosomes showing a small chromosome composed entirely of *S. pombe* DNA. DNA is stained with propidium iodide and *S. pombe* DNA is highlighted by FISH and indicated by yellow arrows. Scale bars indicate 5 µm (in **A**) or 1 µm (in **B**).

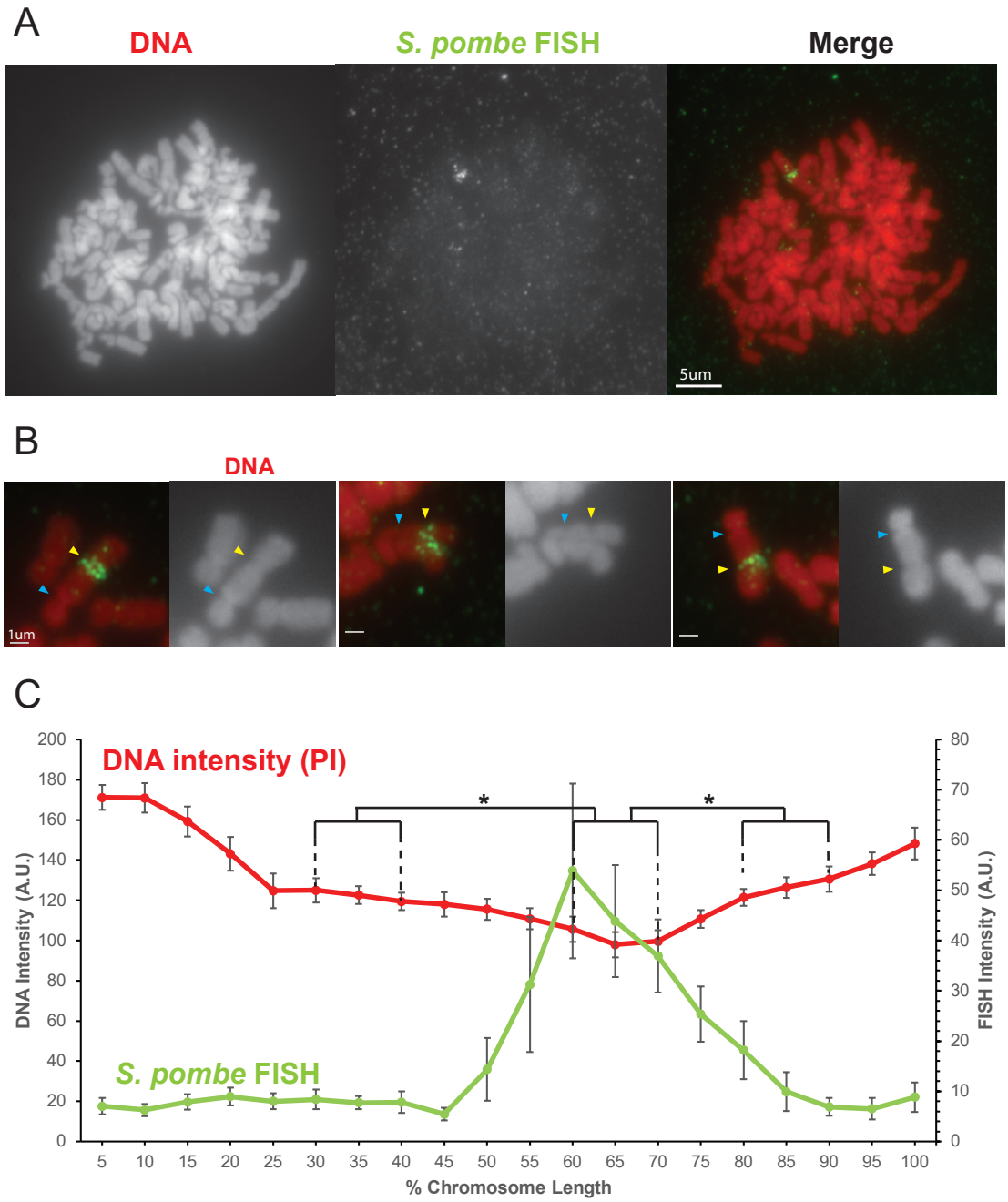


**Figure 4.8 - A distinct chromatin structure forms on *S. pombe* DNA inserted in a human HeLa cell chromosome**

**A & B** - Metaphase spreads of HeP-3 chromosomes showing the insertion of a large amount of *S. pombe* DNA with a distinct structure. DNA is stained with propidium iodide and *S. pombe* DNA is highlighted by FISH and indicated by yellow arrows. Centromeres appear as constrictions along the length of the chromosome and are indicated by blue arrows. Scale bars indicate 5 µm (in **A**) or 1 µm (in **B**).
**C** - Average chromosome profile of the HeP-3 insert bearing chromosome across several images (n=15). Signal intensities of DNA stain (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average DNA stain intensity was compared between the region of 60-70% (*S. pombe* DNA corresponding to the highest FISH signal) and either 30-40% or 80-90% (endogenous mouse DNA) by the KS test (*p<0.01).

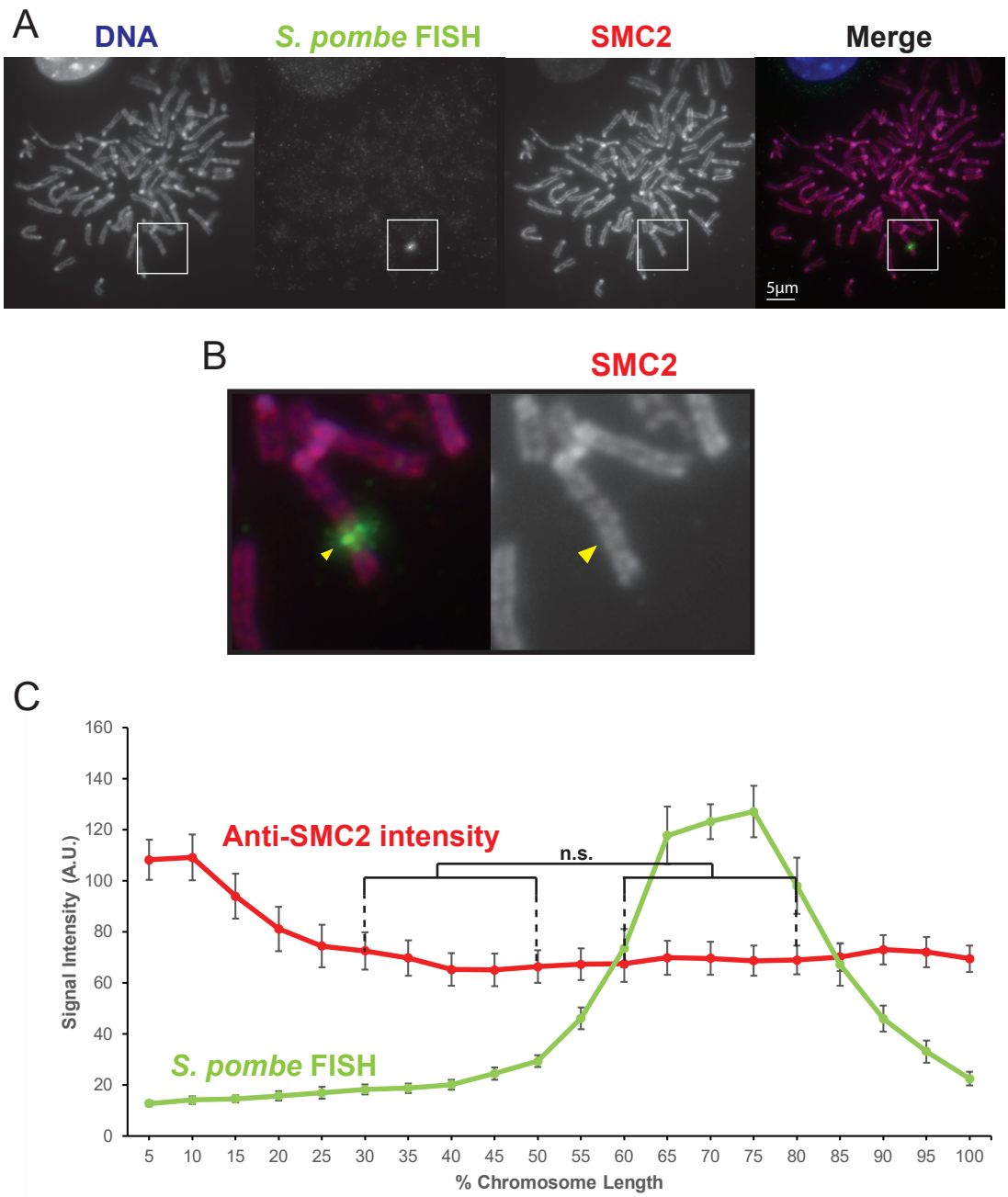


**Figure 4.9 - SMC2 levels show no variation over the site of *S. pombe* DNA insertion**

**A** - Immunolocalisation of F1.1 metaphase spreads for the condensin subunit SMC2 (red channel) with FISH for *S. pombe* DNA (green channel). DNA is stained with DAPI (blue channel). Scale bar indicates 5 µm. The insert-bearing chromosome is highlighted in **A** and shown enlarged in **B**, with yellow arrows indicating the site of *S. pombe* DNA.
**C** - Average chromosome profile of the F1.1 insert bearing chromosome across several images (n=15). Signal intensities of SMC2 IF (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average SMC2 intensity was compared between the regions of 30-50% (endogenous mouse DNA) and 60-80% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01, n.s. = not significant).

The ratio of condensin I to condensin II has been suggested to affect the structure of the mitotic chromosome (Shintomi and Hirano, 2011). A higher condensin I to condensin II ratio was associated with longer, thinner chromosomes, while the opposite proportion resulted in shorter, thicker chromosomes. To determine if there are differences in the ratio of condensin I to condensin II across the *S. pombe* insert, immunolocalisation was performed on the condensin I specific subunit CAPH1 (**Figure 4.10**). The profile of CAPH1 along the insert-bearing chromosome 10 was highly similar to that of overall condensin as assessed by SMC2 levels, with no significant variation in intensity along the entire chromosome arm. These measurements indicate that the ratio of condensin I to condensin II is constant throughout the insert-bearing chromosome 10 for both endogenous mouse and foreign *S. pombe* DNA.

These results suggest that the observed structure of the insert is not explained by any alteration of the organising protein scaffold. Instead, the unusual structure of the *S. pombe* insert must be a result of how the chromatin is organised around this scaffold. Given that the DNA and histone levels decrease across the *S. pombe* insert while the condensin proteins remain stable, the interaction between the chromatin fibres and organising proteins must be altered. This change in interaction can be visualised by observing the ratio of condensin to chromatin. Normalising the previously measured SMC2 and CAPH1 levels to histone H2B levels shows that the ratio of condensin to histones increases significantly at the site of *S. pombe* DNA insertion (**Figure 4.11**). Thus, although the amount of condensin does not change when measured as a function of chromosome length, it does change relative to the amount of chromatin, increasing over the site of *S. pombe* chromatin. This presents several possible interpretations in relation to different models of chromatin folding which will be discussed later.

## 4.5 Cohesin levels increase at the region of foreign chromatin insertion

In addition to condensin, cohesin can mediate chromatin interactions. Immunolocalisation of the cohesin subunit SMC3 was performed on F1.1 metaphase spreads and its levels measured along the length of the insert-bearing chromosome 10 (**Figure 4.12**). Interestingly, unlike condensin, the levels of SMC3

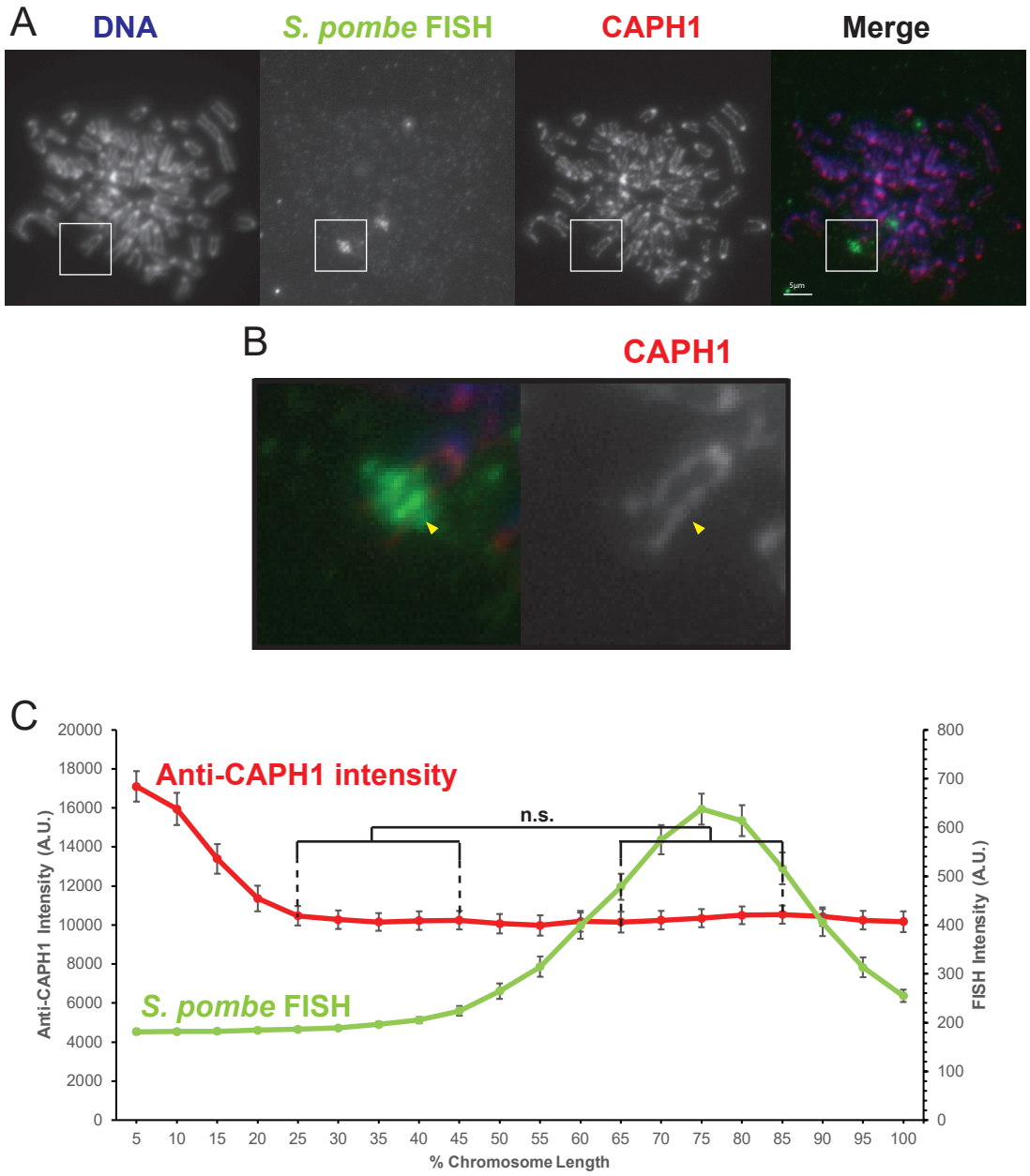


**Figure 4.10 - Condensin I levels show no variation over the site of *S. pombe* DNA insertion**

**A & B** - Immunolocalisation of F1.1 metaphase spreads for the condensin I subunit CAPH1 (red channel) with FISH for *S. pombe* DNA (green channel). DNA is stained with DAPI (blue channel). Scale bar indicates 5 µm. The insert-bearing chromosome is highlighted in **A** and shown enlarged in **B**, with yellow arrows indicating the site of *S. pombe* DNA.
**C** - Average chromosome profile of the F1.1 insert bearing chromosome across several images (n=40). Signal intensities of SMC2 IF (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average CAPH1 intensity was compared between the regions of 30-50% (endogenous mouse DNA) and 60-80% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01, n.s. = not significant).

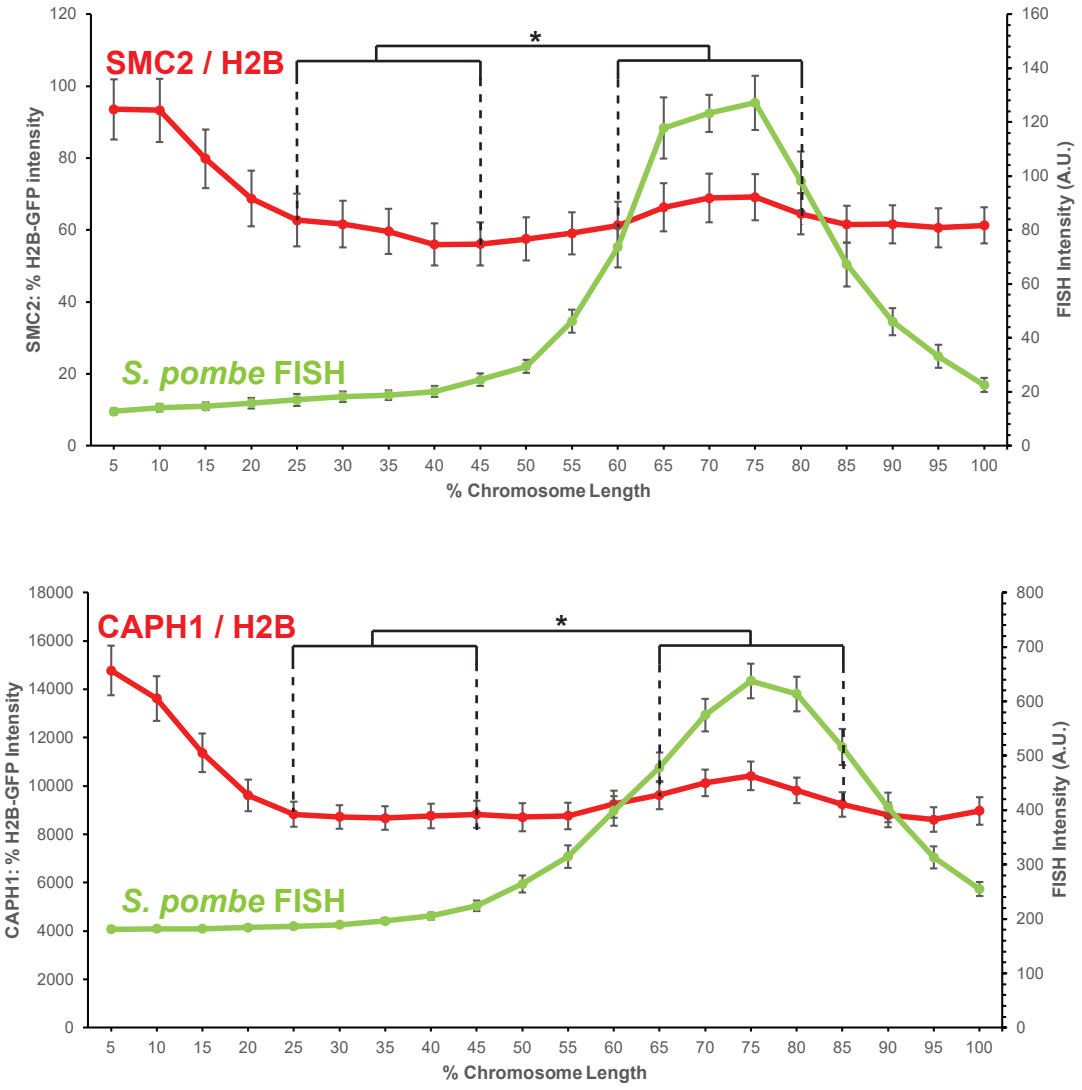


**Figure 4.11 - Condensin levels increase relative to chromatin at the site of *S. pombe* DNA insertion**

Normalisation of the overall condensin (top) and condensin I (bottom) profiles to chromatin levels (as measured by H2B-GFP fluorescence intensity) across the F1.1 insert-bearing chromosome in metaphase. FISH intensities remain unaltered. Error bars represent +/- SEM. Intensity ratios were compared between the control region of 25-45% and the region of *S. pombe* DNA corresponding to the region of highest FISH (either 60-80% or 65-85%) using the KS test (*p<0.01).

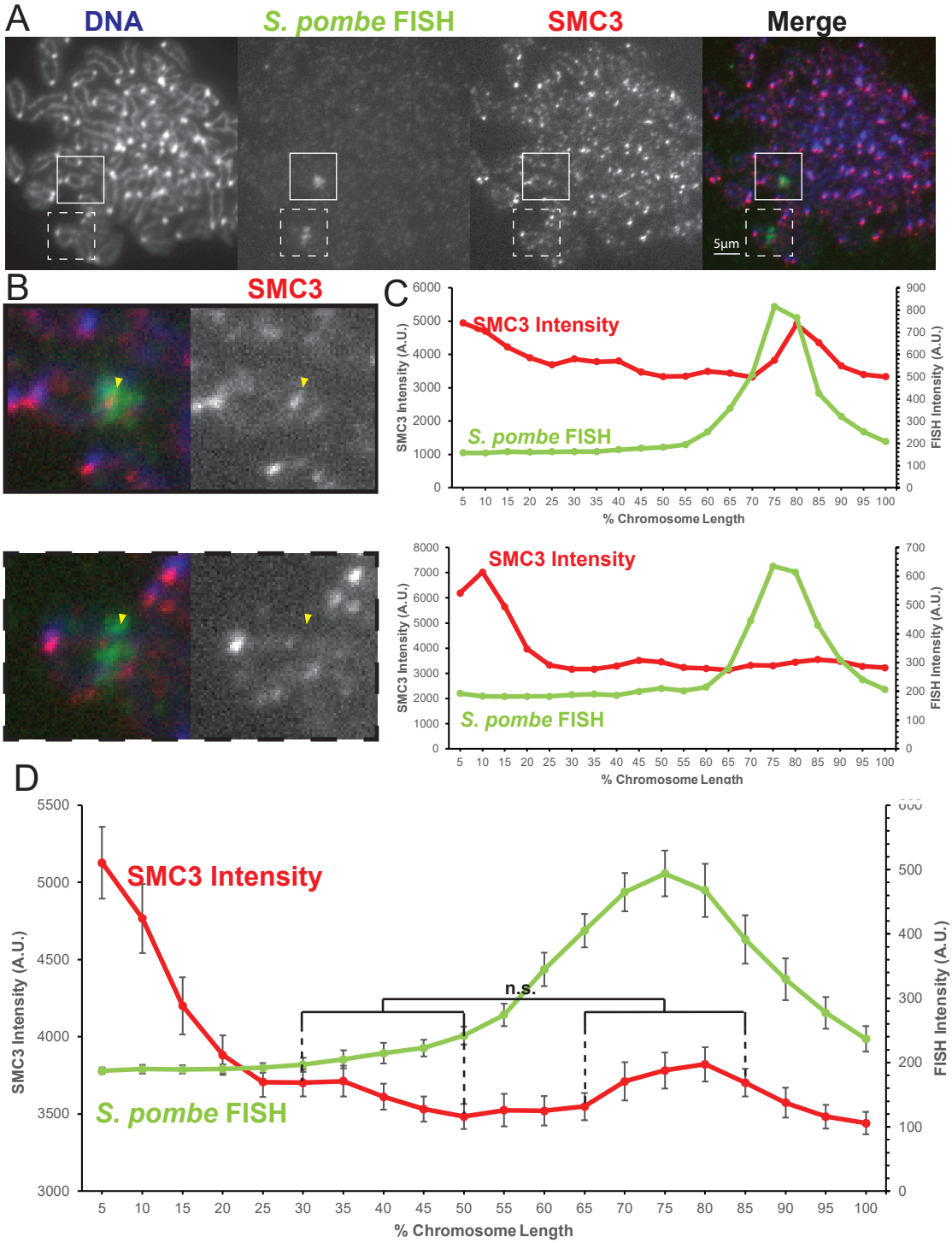


**Figure 4.12 - Cohesin persists over the region of *S. pombe* DNA in metaphase**

**A & B** - Immunolocalisation of F1.1 metaphase spreads for the cohesin subunit SMC3 (red channel) with FISH for S. pombe DNA (green channel). DNA is stained with DAPI (blue channel). Scale bar indicates 5 µm. Two insert-bearing chromosomes are highlighted in **A** and shown enlarged in **B**, with yellow arrows indicating the site of *S. pombe* DNA.
**C & D** - Individual (**C**) and average (**D**) profiles of the F1.1 insert bearing chromosome. Signal intensities of SMC2 IF (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position before plotting. Individual profiles in **C** correspond to the images in **B**. In **D** n=24 and error bars represent +/- SEM. Average CAPH1 intensity was compared between the regions of 30-50% (endogenous mouse DNA) and 65-85% (*S. pombe* DNA) by the KS test (*p<0.01, n.s. = not significant).

showed a small but marked increase in intensity at the site of *S. pombe* DNA insertion compared with the surrounding mouse chromatin on chromosome 10. However, the levels of SMC3 associated with the *S. pombe* chromatin remained considerably lower than the high levels associated with the mouse centromere, and the perceptible increase in SMC3 levels over the *S. pombe* DNA was not significant compared to the rest of the chromosome arm.

It is interesting to note however that in the previous chromosome profiles presented here, the mean intensity levels were quite representative of the individual chromosome profiles collected from specific images, with some attenuation and broadening of the peaks due to the averaging. In the case of SMC3 however, most profiles instead fall into two categories: those in which the insert region bears no distinguishing characteristics and resembles surrounding chromatin, and those in which the FISH signal coincides with a quite stark peak of SMC3 intensity, comparable to the levels at centromeres (**Figure 4.12 B & C**). The mean intensity profile is thus somewhat misleading, and the levels of cohesin at the insert might be more faithfully described by sorting the profiles into categories, with around 2/3 of the chromosomes showing a cohesin signal at the insert and 1/3 showing none.

## 4.6 Hi-C structural analysis reveals a reduced range of interaction for the foreign chromatin in mitosis

In recent years, detailed Hi-C analysis has provided a better understanding of the structure of mitotic chromatin. Analysis of cells throughout the cell cycle has revealed that topologically associated domains (TADs) are dismantled prior to mitosis in human cells (Naumova *et al.* 2013). The characteristic patchwork pattern of intrachromosomal contacts is replaced by a more uniform pattern for all chromosomes with high contact frequencies close to the diagonal and a sharp cut-off threshold beyond which contact frequencies rapidly decline. This pattern is consistent with our understanding of mitotic chromosome organisation. In mitosis we expect a given region of chromatin to be in close proximity (and thus interacting strongly) with adjacent regions of chromatin on the linear mitotic chromosome. However, little or no interaction would be expected between distant regions on the chromosome which are confined to separate domains. In the Hi-C data the observed cut-off threshold is thus interpreted as an indicator of the average size of the

chromatin loops, which is estimated to be between 80 and 120kb. While in this data the loop size remains uniform, in theory a threshold closer to the diagonal would indicate a smaller loop size than a threshold located further out from the diagonal.

In F1.1 we observed an unchanging protein scaffold throughout the insert-bearing chromosome 10, as indicated by constant condensin levels. However, the decrease in chromatin associated with the insert suggests an altered organisation of chromatin around the scaffold which may be detectable by Hi-C. An organisation of the *S. pombe* chromatin into smaller loops, and thus more frequent attachment between chromatin and the axial scaffold, might explain the observed structure. Such an arrangement should be seen by Hi-C as a decrease in the distance of the cut-off threshold for intrachromosomal interactions.

In order to determine the conformation of the *S. pombe* insert chromatin by Hi-C, F1.1 cells were subjected to the same analysis as in this study. Hi-C was performed on both asynchronous (and thus primarily interphase) and synchronised metaphase F1.1 cells, and contact frequencies between chromatin regions were mapped. For the purposes of this analysis contacts between mouse and *S. pombe* chromatin were ignored and mouse-mouse or *S. pombe-S. pombe* contacts were mapped separately to their respective reference genomes.

As previously observed the pattern of TADs of the mouse chromatin in interphase was seen to disappear in favour of a more uniform pattern in mitosis (chromosome 10 is shown as an example in **Figure 4.13**). The overall pattern for the organisation of the insert is more difficult to assess due to the rearrangement of the *S. pombe* DNA. Given that a complete map of the insert was unavailable it was not possible to say which regions of the *S. pombe* genome were adjacent within the insert. It was therefore necessary to analyse the data only within long stretches of DNA whose entire sequence is present in the F1.1 cell line, and is thus likely to be contiguous. For this, two regions of *S. pombe* DNA were chosen. The first was a region of 680kb from *S. pombe* chromosome III, between coordinates 1,720,000 and 2,400,000. This region has the benefit of being one of the largest stretches of contiguous *S.pombe* sequence in the F1.1 genome. It also shows little variation in copy number, suggesting that there is little or no rearrangement within the region. However, the number of sequencing reads over this region suggests that it is present in two

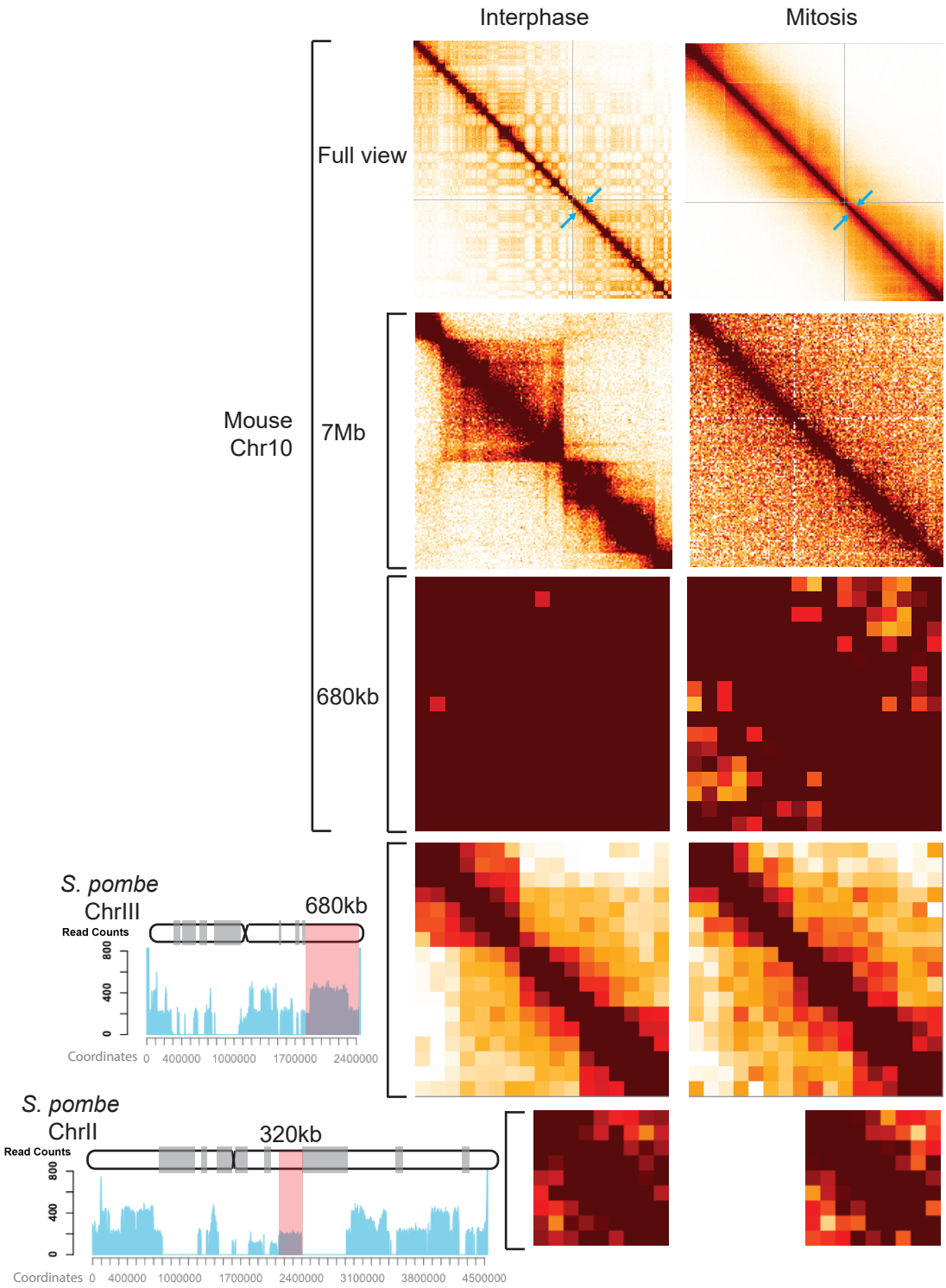


**Figure 4.13 - Hi-C interaction patterns suggest a small-loop organisation of F1.1 insert chromatin in mitosis**

Hi-C interaction heatmaps of identical regions of F1.1 chromatin in interphase (left) and metaphase (right). For the whole chromosome map each pixel represents 250kb. For the other smaller regions each pixel represents 40kb. Blue arrows indicate the site of *S. pombe* DNA insertion. The position of the regions of *S. pombe* sequence under examination within the *S. pombe* genome are highlighted in red to the right on both the diagram and the F1.1 sequencing read maps previously shown in **Figure 3.1**.

copies within the F1.1 *S. pombe* insert, which may complicate analysis. To compensate for this shortcoming, a second smaller region of 320kb from *S. pombe* chromosome II, between coordinates 2,080,000 and 2,400,000, was also examined. While this region's smaller size made it less convenient for analysis, the entirety of the region was present at only one copy within the F1.1 genome, eliminating the possibility of confounding factors from a second copy as well as making it highly likely that the region is contiguous and unrearranged within the *S. pombe* insert. Both regions were located on one of the arms of their respective *S. pombe* chromosomes, and thus did not contain any centromeric or pericentromeric sequences which might complicate the analysis or interpretation of the Hi-C data.

The average mouse TAD size is 1Mb, making it difficult to draw conclusions concerning the interphase organisation of the *S. pombe* insert from regions of only a few hundred kb. However, it is notable that some substructure is visible within these *S. pombe* regions in asynchronous F1.1 cells, while regions of similar size in mouse chromatin show near-saturation of the interaction frequency across the region. This could indicate that the *S. pombe* chromatin is organised on a different scale to the mouse chromatin in interphase as well as mitosis.

The change in the pattern between asynchronous cells and mitotic cells is less pronounced in the *S. pombe* regions than in the mouse regions, although the larger 680kb region of *S. pombe* chromatin does seem to lose some of its compartmentalisation in favour of a more uniform diagonal interaction pattern, just as in mouse chromatin. In contrast to mouse chromatin however, both regions of *S. pombe* insert chromatin show an interaction pattern much more tightly localised to the diagonal. Indeed, the interaction frequency for the mouse chromatin remains saturated within 300 to 400kb of the diagonal, before declining sharply. For the *S. pombe* regions, the decline in frequency occurs at a distance of 120 to 160kb from the diagonal.

The Hi-C data therefore suggests that the *S. pombe* insert chromatin in the F1.1 cell line is organised into smaller loops than the mouse chromatin in mitosis. This altered organisation results in a thinner appearance of the foreign *S. pombe* chromatin on the mitotic chromosome, as observed by microscopy.

## 4.7 Discussion

The results presented here confirmed that the recovered F1.1 cell line retains a distinct structure over the *S. pombe* insert as originally described (**Figure 4.1**, McManus *et al.*). Propidium iodide was used as a DNA stain both in order to replicate the original experiment, and due to its lack of DNA sequence preference (Martin *et al.* 2005). The more commonly used DAPI has a well-known A/T preference, which could lead to differential staining based on sequence composition (Kubista *et al.* 1987, Martin *et al.* 2005). Given the overall greater A/T richness of the *S. pombe* genome compared to the mouse genome (Ruvinsky and Graves 2005, Wood *et al.* 2002), PI staining provides an unbiased and therefore better indicator of fluorescence intensity, which can then be directly related to the amount of DNA along the length of a chromosome.

The generation of additional mouse (NP-A4) and human (HeP-3) cell lines containing large blocks of *S. pombe* chromatin with a similar constricted mitotic appearance showed that this unusual structure is a general property of *S. pombe* chromatin inserted into mammalian chromosomes (**Figures 4.6 & 4.8**). This raises the question of what exactly is informing the processes that recognises this chromatin as being different and leads to its different structure. A primary candidate would be sequence features within the *S. pombe* DNA itself. However, there are several arguments against a purely DNA sequence based process. First, the *S. pombe* DNA sequences present in the three cell lines examined (F1.1, NP-A4 and HeP3) are likely to be quite different in composition, making it unlikely that particular sequences are important. This however still leaves the possibility that more general features of the sequence, such as base composition, are involved. The A/T richness of the *S. pombe* genome is higher than that of mouse (64% compared to 58%) (Ruvinsky and Graves 2005, Wood *et al.* 2002), which could affect the chromatin structure. Such an explanation would be in accordance with older concepts of chromatin organisation around A/T rich scaffold attachment regions (Adachi *et al.* 1989), but these models are no longer widely accepted.

A second potential cause of the unusual chromatin structure could be chromatin modifications. The *S. pombe* DNA introduced into the mammalian host cell is already fully assembled on chromatin and will carry with it all of the chromatin marks

it possessed in the original donor cell. The recognition of these pre-existing histone modifications on *S. pombe* chromatin by the mammalian proteins could lead to propagation or establishment of new domains of histone modifications, which could in turn lead to a distinct structure of the chromatin. Determining the pattern of histone modifications over the *S. pombe* insert is thus an important step in understanding the factors that determine its structure.

In order to increase our understanding of the nature of the structure at the F1.1 *S. pombe* insert, examination of the insert chromatin by methods other than DNA staining was necessary. Microscopy of two separate core histones visualised by two separate methods confirmed that chromatin levels do indeed decrease at the foreign DNA insert (**Figures 4.4 & 4.5**). These results therefore rejected the possibility of an increase in chromatin compaction explaining the distinct structure, which would be expected to increase the levels of chromatin in a given area as the chromatin was condensed into a smaller volume than endogenous chromatin. This decrease in chromatin levels was not mirrored by condensin levels, which remained constant across the whole length of the chromosome (**Figures 4.9 & 4.10**). These results suggested a model whereby the foreign chromatin is organised into smaller loops around an unchanging protein scaffold of condensins. This model would explain both the narrower diameter of the mitotic chromosome previously observed by EM (McManus *et al.*) and the decrease in chromatin levels. Hi-C analysis within regions of *S. pombe* chromatin in F1.1 also supported this model, as the drop-off of interaction frequency indicated a smaller loop size within *S. pombe* chromatin compared with mouse chromatin in mitosis (**Figure 4.13**).

The adopting of a smaller loop size by the *S. pombe* chromatin may be explained in various ways within the radial loop / loop extrusion model of mitotic chromosome structure. One possible explanation is that a localised alteration of the ratio of condensin I to condensin II across the *S. pombe* insert mediates a change in chromosome architecture. Such a change in condensin I to II ratio over whole chromosomes has been observed to alter chromosome shape (Shintomi and Hirano 2011). However, no variation in the condensin I to II ratio was detected over the *S. pombe* insert. Alternatively, a change in the activity of condensin relative to its DNA substrate could lead to a change in structure without a change in ratio if it affected

both condensin complexes equally. This change could primarily result from the alteration of two parameters of condensin activity (**Figure 4.14**):

1. An increase in the binding or loading of total condensin. If we assume that the alignment and settling of the condensin into an axial scaffold results from the interaction between condensin complexes, as has been suggested (Goloborodko *et al.* 2016, Gibcus *et al.* 2018), then it is to be expected that overall levels of condensin should remain stable along the chromosome axis. Increased loading of condensin would therefore not result in an increase in condensin levels on the axis but rather a decrease in DNA around the axis, just as we observe at the *S. pombe* insert. The greater number of condensin complexes per length of chromatin means that they would encounter each other sooner as they extruded their loops of chromatin, blocking further extrusion and settling into a small-loop configuration.

2. A decrease in the processivity of condensin. Condensin complexes continue to bind the chromatin fibre while other condensins that were previously bound continue to extrude chromatin loops. New condensin can either bind within an existing loop, in which case it forms a loop within a loop, or outside an existing loop, in which case it forms a new loop that will extrude until it is hindered by an adjacent condensin complex. If the processivity of condensin decreases, for instance if it extrudes chromatin at a slower rate, then the probability of forming side-by-side loops increases relative to the probability of forming loops-within-loops. A decrease in processivity thus also results in a small-loop configuration.

A local change in one or both of these parameters of condensin at the *S. pombe* insert could therefore alter the mitotic chromosome structure compared with the surrounding chromatin. However, the conditions that might alter condensin behaviour in this way remain unknown.

It also remains unclear whether cohesin is contributing to the unusual structure of the *S. pombe* insert. Cohesin levels increased over the *S. pombe* insert in around two-thirds of F1.1 metaphase spreads (**Figure 4.12**). Rather than representing an increase in loading, these results may thus instead be describing a persistence of cohesin in the region. This would be consistent with previous observations of a small

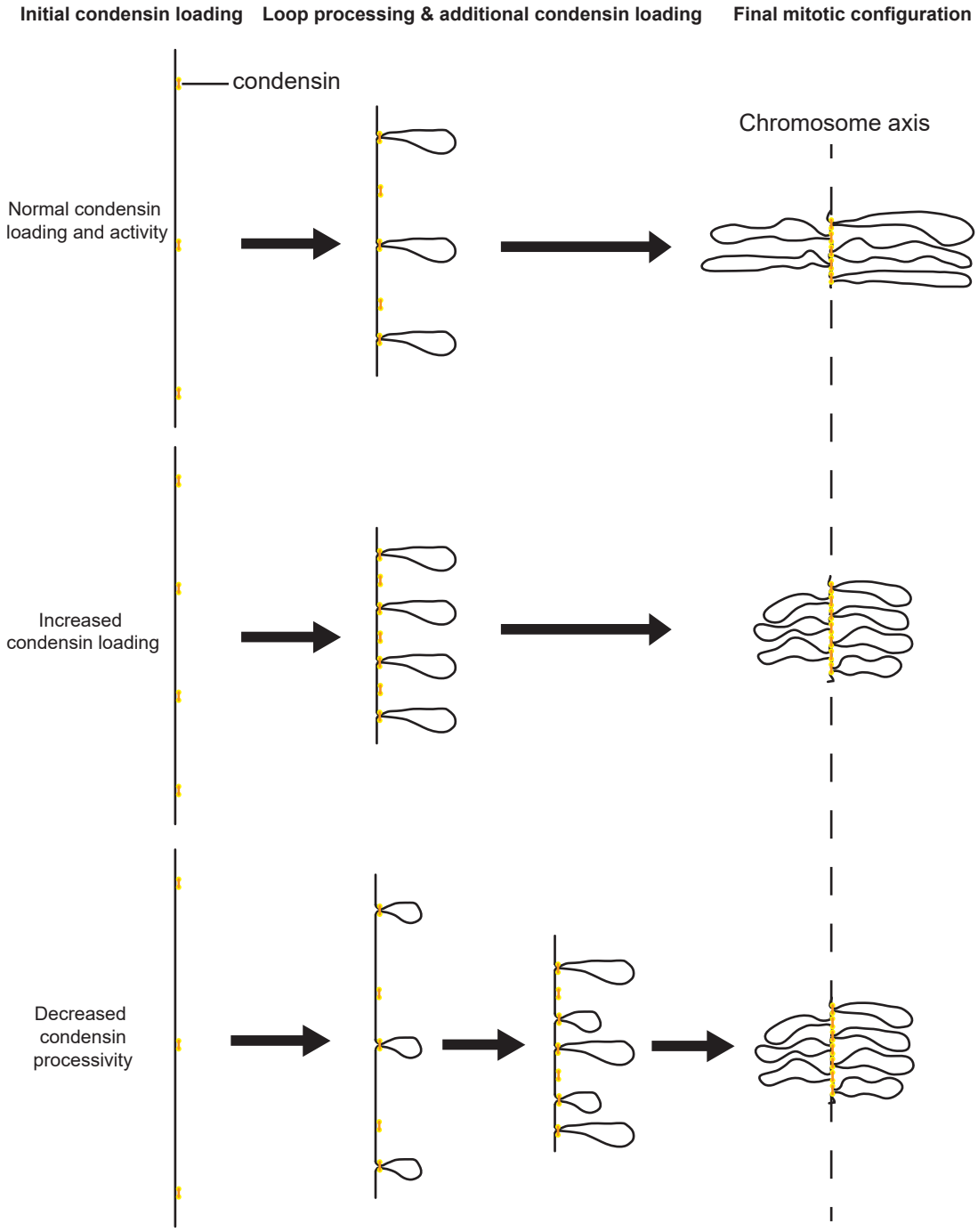


**Figure 4.14 - Organisation of chromatin into smaller loops by altering condensin activity**

Diagram of chromatin organisation in mitosis according to the radial loop / loop extrusion model. Loops-within-loops are ommitted for clarity. Interaction between condensins results in the same amount of condensin along the chromosome axis in all cases, however changing the condensin parameters results in a different organisation of chromatin into loops. Normal condensin loading and processivity results in a normal, “large-loop” configuration (Top). If more condensin is loaded over the same size of chromatin fibre, the loop size decreases accordingly (middle). If condensin extrudes loops more slowly, then other condensin complexes have more time to bind between extruding loops, and loop size is also decreased (bottom).

percentage of cells in which the chromatids remain attached at the site of insertion late into the cell cycle (McManus *et al.* 1994). Whether such variation is a cause of, or even related to, variation in structure is another question. Although cohesin has been shown to affect mitotic chromosome structure, it is believed to act indirectly by altering the ratio of condensin I to II (Shintomi and Hirano 2011). We detect no change in this ratio over the *S. pombe* insert, ostensibly ruling out this mechanism. It may be that our methods are not sufficiently sensitive to detect the ratio change caused by the cohesin, or that cohesin itself plays a more direct role in chromatin architecture than previously thought.

A related question is why does the cohesin persist over the *S. pombe* insert at all? Cohesin is usually removed from the chromosome arms early in mitosis, leaving only a small pool at centromeres until anaphase (Waizenegger *et al.* 2000). Cell cycle differences may also explain the increase of cohesin in some cells and not others. The majority of chromosomes imaged are in pro-metaphase or metaphase, but the differences in cell cycle timing, or in total time spent in mitosis, could still account for some variation in the cohesin levels. Another possibility is that the *S. pombe* insert possesses some features in common with centromeric chromatin, thus protecting it from cohesin removal through the same mechanism that protects centromeres. This further raises the possibility that the observed persistence of cohesin is a symptom of some feature of the underlying chromatin structure, rather than a causative agent of the structure itself. Heterochromatin has been found to be a key component of centromeric regions as well as being important in the protection of cohesin from removal (Bernard and Allshire 2002, Yi *et al.* 2018). Cohesin persistence therefore also highlights the importance of determining the chromatin profile of the F1.1 insert.

# Chapter 5: *S. pombe* chromatin inserted in mammalian cells is characterised by a highly repressive chromatin environment

## 5.1 Introduction

The previous chapter examined in more detail the nature of the distinct mitotic structure of *S. pombe* chromatin in mammalian cells. However, it remains to be determined why this structure is localised on the *S. pombe* chromatin. So far, the only aspects of the *S. pombe* chromatin found to be distinctive have been directly related to its structure, such as DNA and histone levels. To investigate the origin of these structural changes, distinguishing chromatin features that are not directly related to structure must first be identified. If a non-structural feature of the chromatin is also found to be distinct at the *S. pombe* insert, the possibility of a link between it and the structure of the insert must then be investigated. This can be achieved by altering the chromatin and observing any effect this may have on chromosome structure. The primary candidates for non-structural features to investigate are chromatin modifications.

The nature of the chromatin modifications of the F1.1 *S. pombe* chromatin was not previously investigated as their role in the regulation of chromatin and the experimental methods to investigate them were not yet fully established. Since then the variety of chromatin modifications and their functions has become evident. In particular, the influence of particular histone modifications on chromatin structure make them primary candidates for components which may mediate the formation of the distinct chromatin structure on the *S. pombe* DNA. However, the structural role of histone modifications is best understood at the lowest level of chromatin structure, the 10nm fibre, while their role in mitotic chromosome structure is not well defined (Kouzarides 2007, Bannister and Kouzarides 2011, Zhiteneva *et al.* 2017). Establishing a firm link between chromatin modifications and mitotic chromatin structure would be an important element towards understanding what factors influence the folding of chromosomes in mitosis.

Therefore, in this chapter the association of several chromatin modifications with *S. pombe* DNA inserted in mammalian cells was investigated to determine if they might be distinct from the surrounding regions of mouse chromatin.

## 5.2 The F1.1 *S. pombe* insert is a large region of constitutive heterochromatin

In order to identify characteristics that distinguish the *S. pombe* chromatin insert region aside from its structure, its chromatin profile was determined by investigating the levels of several histone PTMs on the *S. pombe* chromatin in the F1.1 cell line. ChIP-qPCR was performed on F1.1 cells using qPCR primer pairs for *S. pombe* genes known from the sequencing data to have been inserted in the F1.1 genome and located at various regions across the *S. pombe* genome. These loci were compared to several endogenous mouse control loci.

Strikingly, all loci tested in the insert region were observed to have very high levels of the heterochromatic H3K9me3 mark, comparable to the pericentromeric major satellite positive control regions (**Figure 5.1 A**). At all but one of these loci the H3K9me3 levels were significantly higher than at the euchromatic mouse gene actin. This result was corroborated by the presence of significantly high levels of the heterochromatin-associated HP1α protein at all of the insert loci. To confirm the ChIP results immunolocalisation of H3K9me3 followed by *S. pombe* FISH was performed on F1.1 metaphase spreads (**Figure 5.1 B-D**). In all spreads a strong H3K9me3 signal coincided exactly with the region of *S. pombe* DNA insertion highlighted by FISH.

In mammalian cells H3K9me3 is usually associated with DNA methylation at CpG dinucleotides (5meC). Previous work showed using 5meC-sensitive restriction digestion that 5meC increased on the insert over time and is likely at or near saturation at the current number of passages (McManus *et al.* 1994). To confirm this result, F1.1 genomic DNA was bisulfite converted and ten different *S. pombe* DNA fragments corresponding to genes were sequenced. The results are presented in schematic form in **Figure 5.2**. With the exception of one gene (Omh3) in which all CpGs appeared to be unmethylated, the nine other genes had the majority of their CpGs methylated in all experimental repeats while the rest of the CpGs

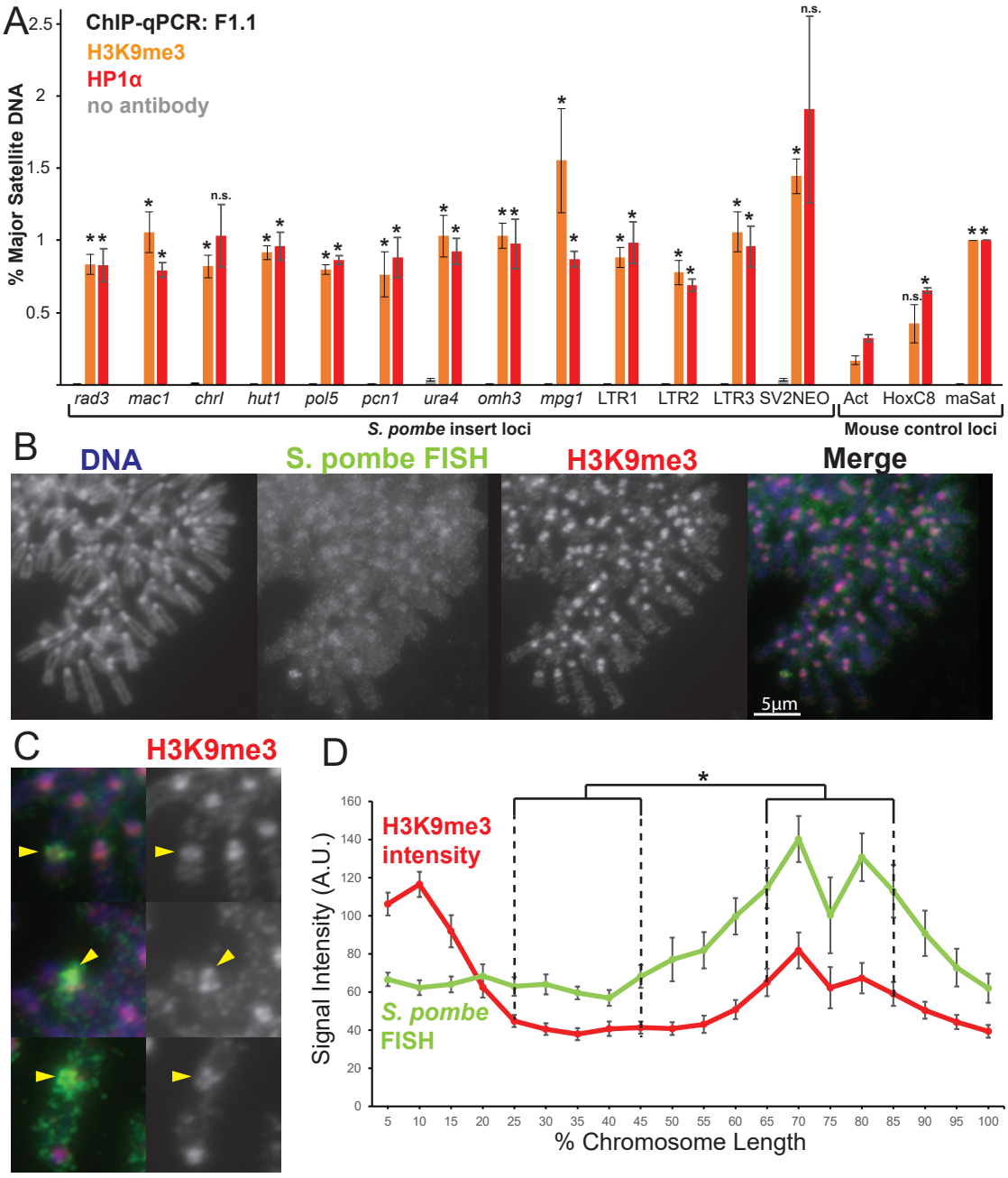


**Figure 5.1 - *S. pombe* chromatin inserted in F1.1 has high levels of H3K9me3 and HP1**

**A** - ChIP-qPCR of F1.1 interphase cells for H3K9me3 and HP1α at 13 loci within the *S. pombe* insert and 3 mouse control loci. Act is the highly transcribed negative control gene actin, HoxC8 is a region of facultative heterochromatin and maSat is the centromeric satellite positive control region. Error bars represent +/- SEM of three independent repeats. Enrichments were normalised to maSat levels and compared to Act by the *t*-test (*p<0.05, n.s. = not significant).
**B & C** - Immunolocalisation for H3K9me3 (red channel) on F1.1 metaphase spreads showing high levels of H3K9me3 at the *S. pombe* insert region as visualised by FISH (green channel). DNA is stained with DAPI (blue channel). Scale bar in **B** indicates 5 μm. The site of *S. pombe* DNA is indicated in **C** by yellow arrows.
**D** - Average chromosome profile for H3K9me3 on the F1.1 insert bearing chromosome across several images (n=14). Immunolocalisation (red) and FISH (green) signal intensity were measured along the length of the chromosomes and binned according to their position. Error bars represent +/- SEM. Average H3K9me3 intensity was compared between the regions of 25-45% (mouse DNA) and 65-85% (*S. pombe* DNA) by the KS test (*p<0.01).

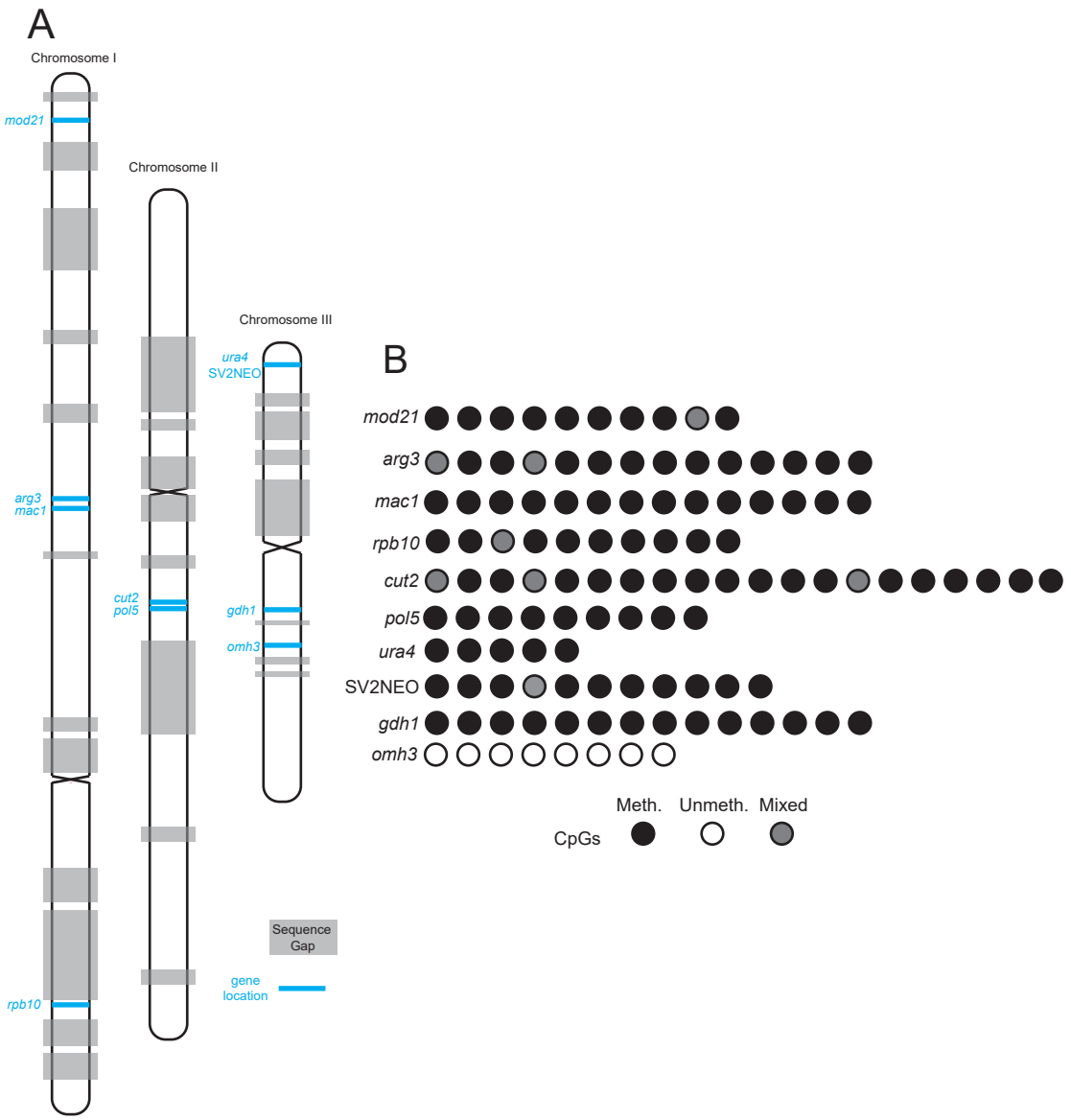


**Figure 5.2 - *S. pombe* DNA inserted in F1.1 is highly CpG methylated**

**A** - Schematic representation of the DNA content of the F1.1 *S. pombe* insert, as determined in **Figure 3.1**. Regions found to be present in F1.1 are white while those found to be absent are greyed out. Location of genes analysed by bisulfite sequencing in **B** are indicated by blue bars.
**B** -CpG methylation levels at ten *S. pombe* loci within the F1.1 insert as determined by bisulfite conversion of F1.1 genomic DNA followed by sequencing in three replicates. Each circle represents a CpG dinucleotide with its methylation state indicated by its colour, corresponding to CpGs found to be methylated in all repeats (black), in only one or two of the repeats (grey), or unmethylated in all repeats (white).

(approximately 10% of the total) were methylated in at least one of the repeats. These results therefore also suggest that high levels of 5meC are present over the *S. pombe* DNA insert in F1.1 cells.

ChIP for other histone marks usually associated with active chromatin was also performed, including histone H3 acetylation at H3K9 and H3K27 and the activating methylation marks H3K4me3 and H3K36me3. As would be expected in a transcriptionally repressed heterochromatic region, and again comparable to pericentromeric chromatin, levels of all these activating histone marks were very low at all regions within the insert tested (**Figure 5.3**). Together, these analyses point to the foreign chromatin having adopted the profile of constitutive heterochromatin, similar to centromeric regions.

## 5.3 The heterochromatic chromatin signature is confined to the region of *S. pombe* DNA

H3K9me3 is known to be capable of spreading over adjacent regions of chromatin unless otherwise checked (Allshire and Madhani 2017). The immunolocalisation in **Figure 5.1** shows a tight coincidence between FISH and anti-H3K9me3 signals, suggesting that the H3K9me3 may not have spread far, if at all, outside of the region of *S. pombe* DNA insertion. However, the resolution of microscopy does not allow us to conclude this with accuracy. To investigate this in more detail, further ChIP-qPCR analysis for H3K9me3 was performed, focusing on the regions of DNA immediately surrounding the site of insertion as identified in Chapter 3. qPCR primers were designed for a region of DNA only a few hundred base pairs away from the site of insertion on mouse chromosome 10, as well as regions on either side at a distance of 1, 2, 10, 100 and 500 kb. H3K9me3 levels at these loci were analysed and compared to the same regions in the parental C127 cell line (**Figure 5.4**). Given that the C127 and F1.1 cell lines are tetraploid, there are three other copies of chromosome 10 per cell aside from the insert-bearing chromosome. We would therefore not expect a large change in H3K9me3 levels if only one copy of chromosome 10 had an increase in H3K9me3. Nevertheless, the H3K9me3 levels between F1.1 and C127 were strikingly similar at all loci examined and showed no significant difference, even at those regions closest to the heterochromatic foreign DNA. Thus, though the previous ChIP results show that heterochromatin is quite

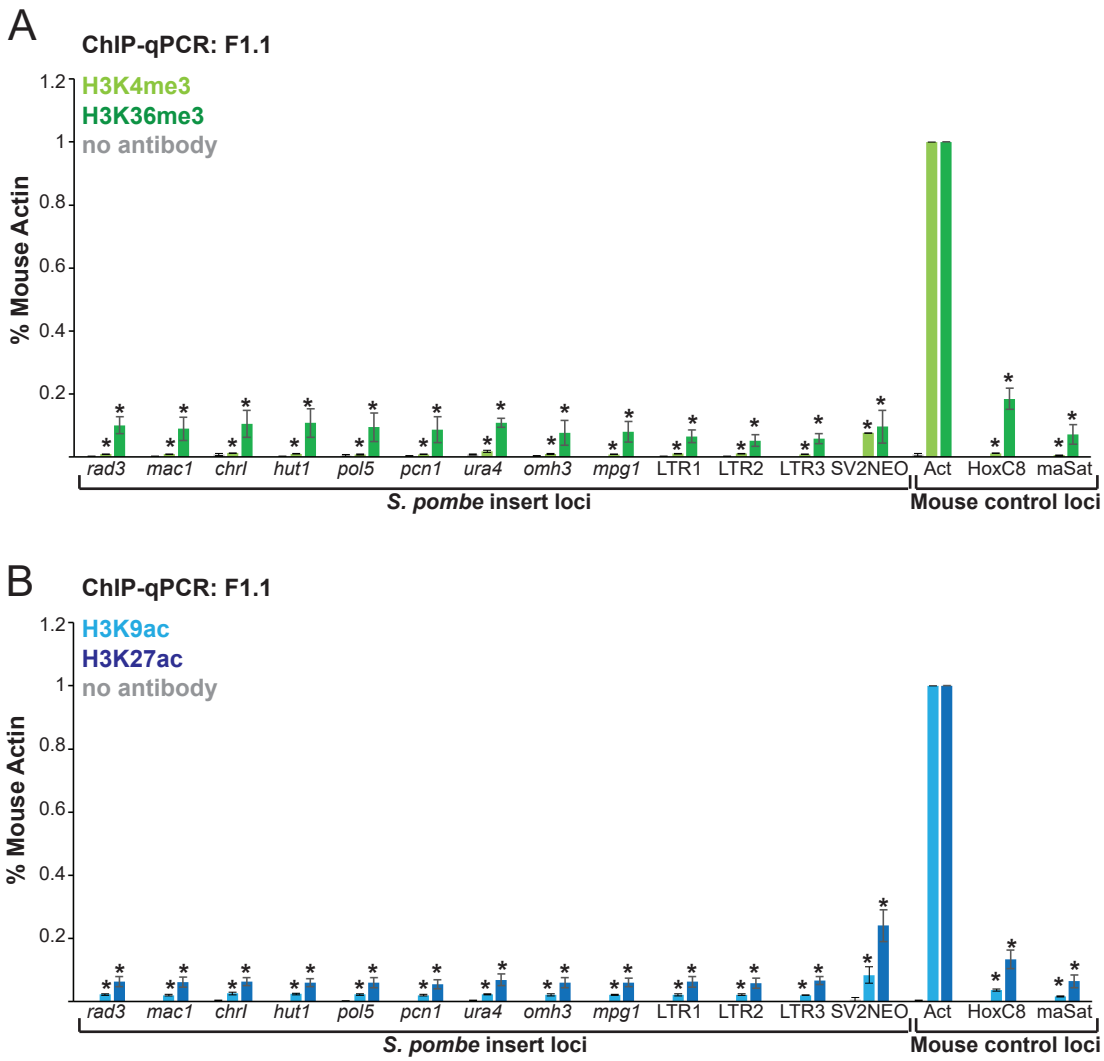


**Figure 5.3 - Activating chromatin marks are absent from the F1.1 *S. pombe* insert chromatin**

ChIP-qPCR of F1.1 interphase cells for activating histone methylation H3K4me3 and H3K36me3 (**A**) and acetylation H3K9ac and H3K27ac (**B**) at 13 loci within the *S. pombe* insert and 3 mouse control loci. Act is the highly transcribed positive control gene actin, and negative control regions are the facultatively heterochromatic HoxC8 gene and the constitutively heterochromatic maSat centromeric satellite repeat. Error bars represent +/- SEM of three independent repeats. Enrichments were normalised to Act levels and compared to Act by the *t*-test (*p<0.05).

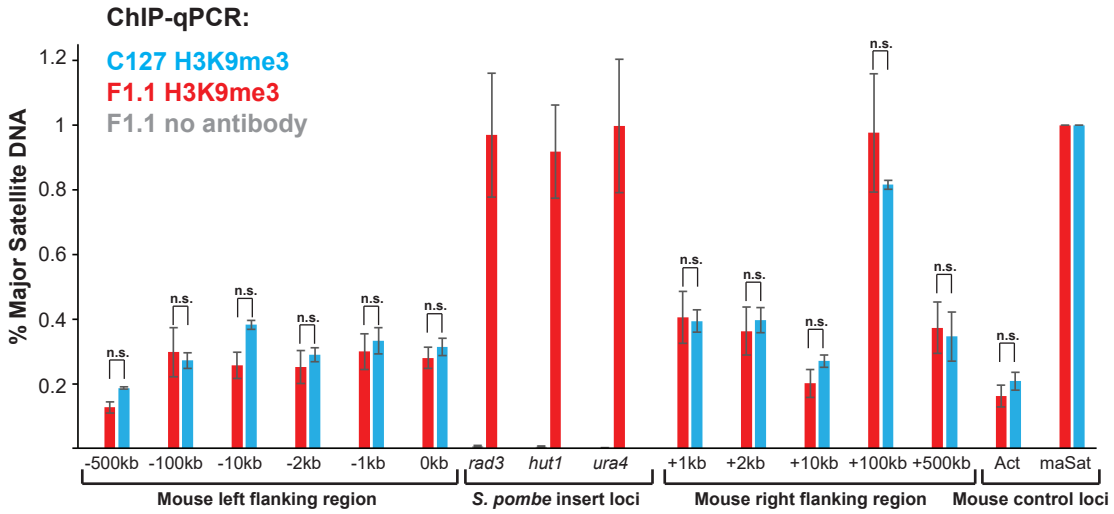


**Figure 5.4 - Heterochromatin is confined to the region of foreign *S. pombe* DNA in F1.1 cells**

ChIP-qPCR for H3K9me3 in F1.1 (red) and C127 (blue) interphase cells at 3 *S. pombe* loci within the insert (for F1.1 only), 2 mouse control loci and 11 regions of mouse DNA on the regions flanking the *S. pombe* chromatin insertion site at the indicated distance to left and right. Act is the highly transcribed negative control gene actin and maSat is the centromeric satellite positive control region. Enrichments are normalised to maSat positive control levels. Error bars represent +/- SEM of three independent repeats. Enrichment levels between F1.1 and C127 cells were compared by a *t*-test (*p<0.05, n.s. = not significant).

widespread, potentially ubiquitous, across the *S. pombe* DNA insert, H3K9me3 is confined entirely to this foreign *S. pombe* chromatin, with no detectable encroachment into the surrounding endogenous mouse chromatin.

## 5.4 Genes within the region of foreign DNA are not transcribed

Although the large evolutionary separation between *M. musculus* and *S. pombe* does pose an obstacle to the efficient functioning of the transcriptional machinery from one species on the sequence of the other, there are nonetheless some conserved and potentially recognisable elements of the eukaryotic transcriptional machinery that might be sufficient to induce some level of transcription at the *S. pombe* genes in mouse cells. The presence of H3K9me3 and HP1α, however, suggests that highly repressive heterochromatin is assembled across the *S. pombe* insert that would prevent most transcription.

To determine whether the *S. pombe* DNA is being transcribed in F1.1 cells, RT-qPCR was performed on a small subset of *S. pombe* genes present within the insert and compared to mouse genes with high, low or no expression expected (Cyclophilin A, cMyc and HoxC8 respectively; **Figure 5.5**). Of the 5 *S. pombe* genes tested, 4 showed no detectable transcription. The *ura4*$^{+}$ gene on the other hand showed a low level of transcription, comparable to the lowly expressed mouse cMyc. This is notable because the G418 resistance gene expressed from the SV40 promoter (SV2NEO), expression of which is actively selected for in the F1.1 cell line by continual G418 treatment, is inserted at the *ura4*$^{+}$ locus in the *S. pombe* genome. The SV2NEO gene itself, which contains a strong promoter, also has detectable expression, though only about 2 to 3-fold more than *ura4*$^{+}$ and mouse cMyc.

Thus, the gene expression and ChIP analyses generally agree. A strongly repressive environment was detected across most of the *S. pombe* insert. However, a small increase in activating marks was observed on the SV2NEO gene, in agreement with a low but detectable level of expression of that gene (**Figure 5.3**). The fact that SV2NEO is expressed is expected, as its expression is necessary for the survival of the cells under G418 selection. However, the low levels of expression and activating marks suggest that the SV2NEO gene is a small island of activated chromatin surrounded by repressive H3K9me-dependent heterochromatin.

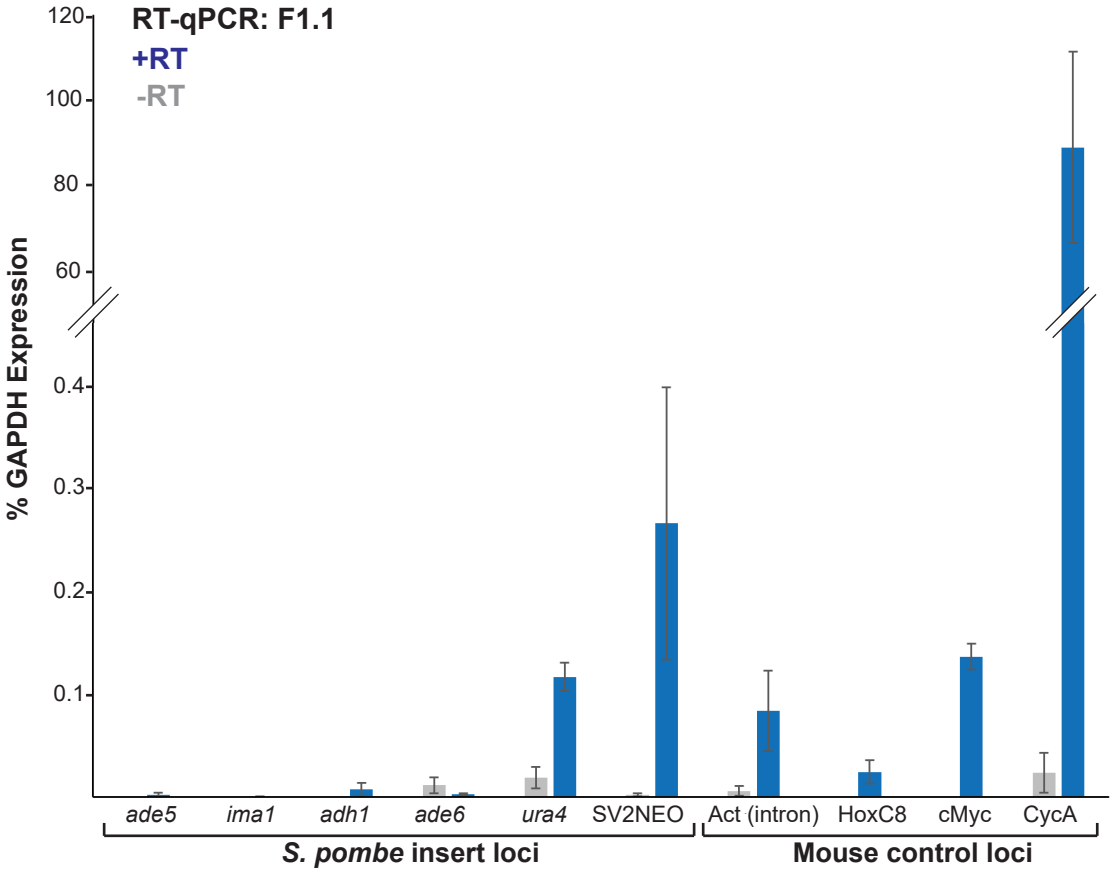


**Figure 5.5 - *S. pombe* genes within the F1.1 DNA insert are not transcribed**

RT-qPCR of total F1.1 RNA. Reactions were incubated with (blue) or without (grey) reverse transcriptase and RNA levels were measure at 5 *S. pombe* loci within the F1.1 insert, as well as the SV2NEO resistance gene. For comparison mouse genes expected to have no expression (HoxC8), low expression (cMyc) and high expresion (CycA) were included, as was an intronic sequence from mouse actin, not expected in final mRNA. Expression levels were normalised to the highly expressed mouse GAPDH gene. Error bars represent +/- SEM of three independent repeats.

## 5.5 A heterochromatic signature is a common feature of *S. pombe* chromatin in yeast-mammalian fusion cells

In addition to the F1.1 cell line, *S. pombe* chromatin was successfully inserted into NIH-3T3 and HeLa cells (see **Chapter 4**). The highly heterochromatic profile of the *S. pombe* chromatin in F1.1 raises the question of whether heterochromatinisation is a common feature of *S. pombe* chromatin inserted into mammalian chromosomes. Since the NP-A4, HeP-2 and HeP-3 cells were generated recently and cryo-preserved after a few passages, the *S. pombe* chromatin will have been present in the host cell for no more than 25 to 30 cell divisions. Thus, examination of the chromatin profile of *S. pombe* chromatin in these cells would determine if heterochromatin is established relatively soon after insertion, or if the F1.1 *S. pombe* insert only gradually assembles heterochromatin over time.

ChIP-qPCR for H3K9me3 was therefore performed on the NIH-3T3 derived NP-A4 cell line and the HeLa-derived HeP2 and HeP3 cell lines (**Figure 5.6**). Whenever possible the same *S. pombe* loci that were analysed in F1.1 were again examined in these cells. However, due to the differing content of the *S. pombe* DNA inserted in each of the cell lines, the repertoire of *S. pombe* loci were not always identical. Nevertheless, all three cell lines were found to have significantly high levels of H3K9me3 at most loci across the *S. pombe* insert, comparable to the centromeric positive control regions in each cell. Thus, as in F1.1, the *S. pombe* chromatin regions inserted into chromosomes of NIH-3T3 and HeLa cells display high H3K9me3 levels consistent with constitutive heterochromatin.

To confirm these results, immunolocalisation of H3K9me3 followed by *S. pombe* FISH was performed on NP-A4 metaphase spreads (**Figure 5.7**). The levels of H3K9me3 over the NP-A4 *S. pombe* DNA did not appear quite as strong as that over the F1.1 *S. pombe* insert. However, this may be due to the fact that the NP-A4 *S. pombe* insert appears to be smaller than the F1.1 insert (compare **Figure 4.1** and **Figure 4.6**). Nevertheless, the H3K9me3 levels on NP-A4 chromosomes did show a significant increase coinciding exactly with the site of *S. pombe* chromatin insertion. While not quite as strong as at centromeres, the H3K9me3 levels over the *S. pombe* insert were still considerably higher than over the rest of the chromatin along the insert-bearing chromosome. These measurements therefore support the ChIP

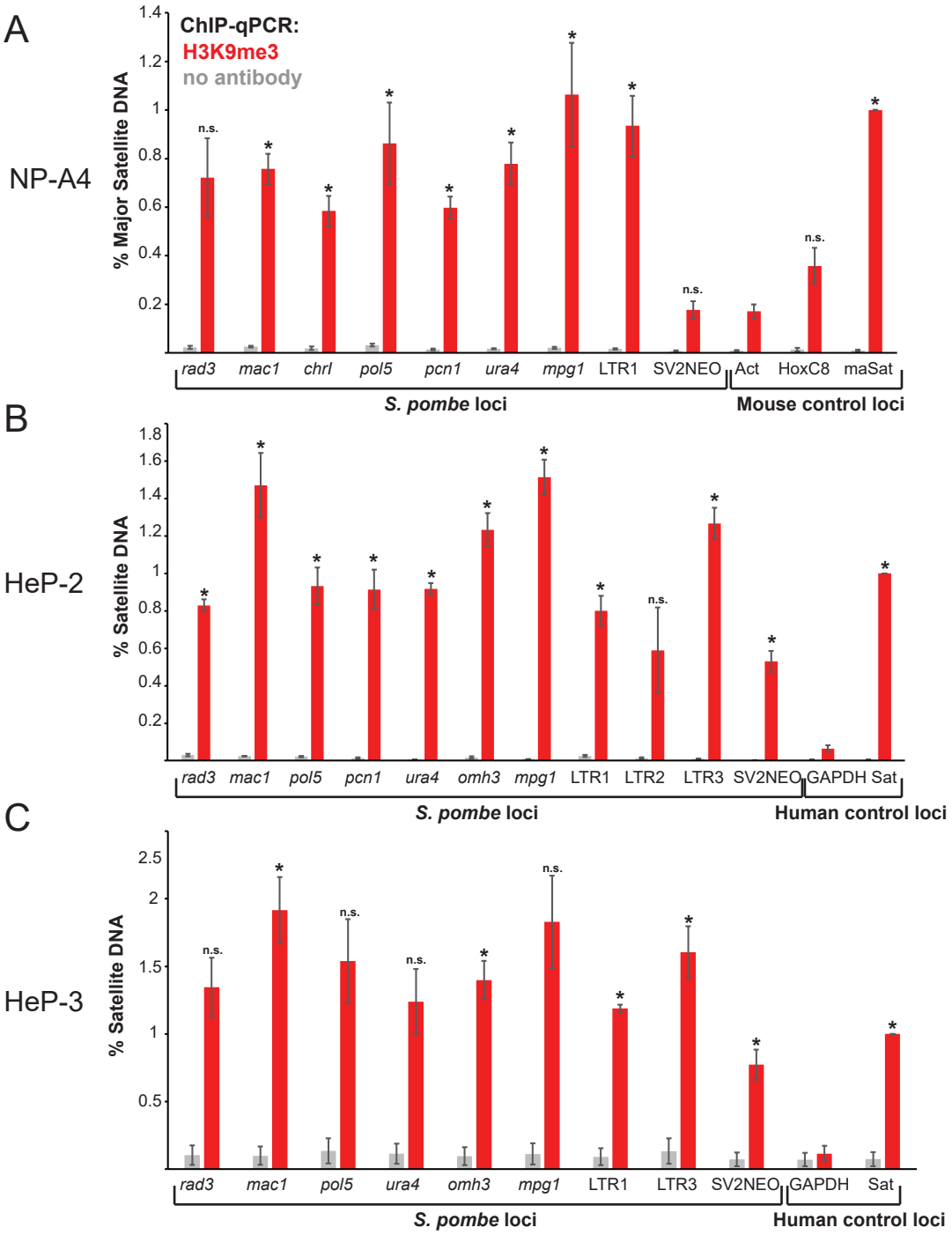


**Figure 5.6 - Heterochromatin is established across *S. pombe* chromatin inserted in several mammalian cell lines**

ChIP-qPCR for H3K9me3 in the mammalian-*S. pombe* fusion cell lines NP-A4 (**A**), HeP-2 (**B**) and HeP-3 (**C**) in interphase. qPCR was performed using primers previously used in F1.1 qPCR. Loci analysed in F1.1 but found not to be present in the new cell lines were omitted from the final analysis. In **A** mouse control regions Act, HoxC8 and maSat were included as before. In **B** and **C** two human control loci were included: the highly transcribed negative control gene GAPDH and a human centromeric satellite DNA positive control region, Sat. Error bars represent +/- SEM of three independent repeats. Enrichments were normalised to positive control levels and compared to the negative control by the *t*-test (*p<0.05, n.s. = not significant).

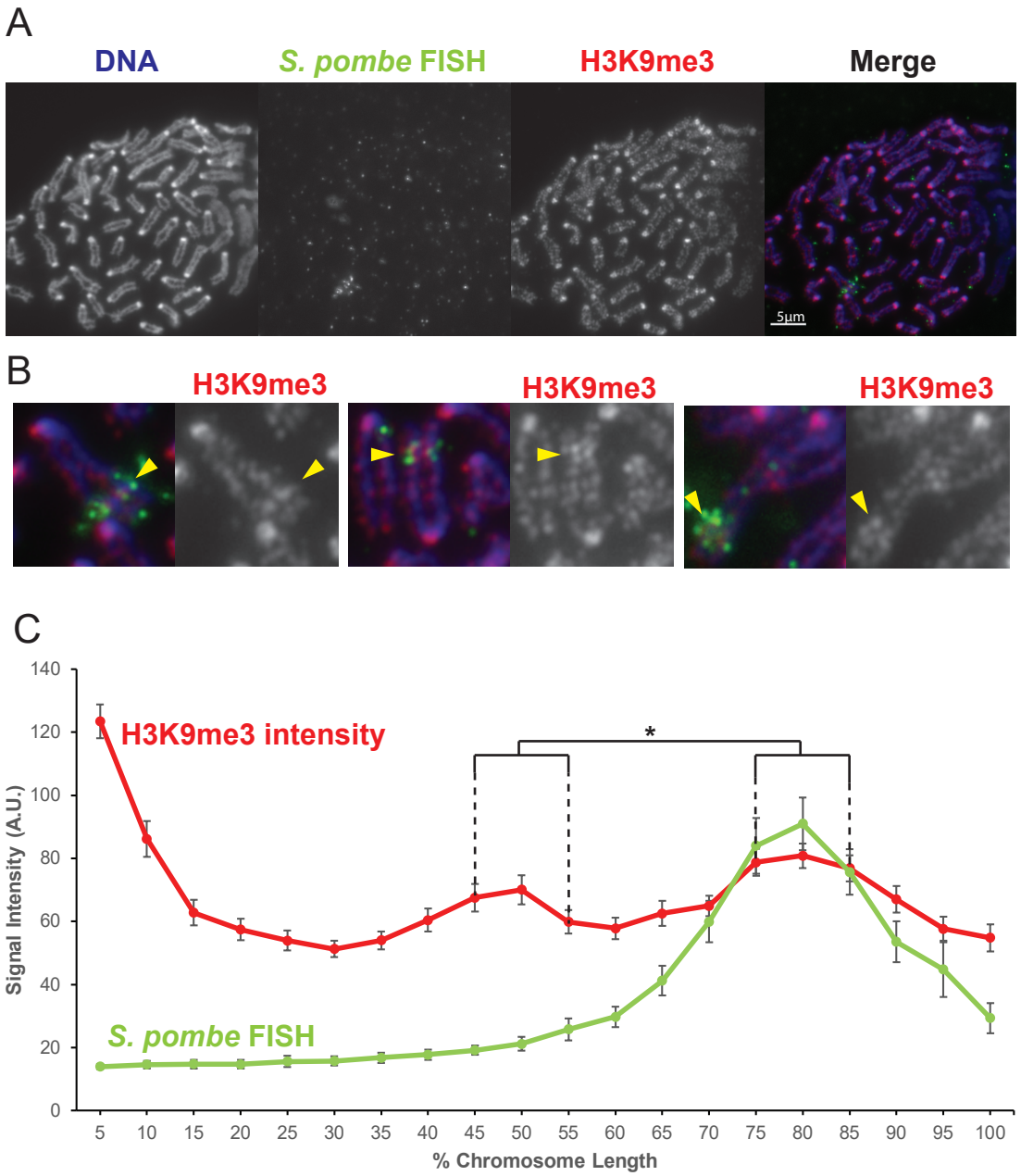


**Figure 5.7 - *S. pombe* chromatin inserted in the NP-A4 cell line has high levels of H3K9me3 visible by immunolocalisation**

**A & B** - Immunolocalisation for H3K9me3 (red channel) on NP-A4 metaphase spreads showing high levels of H3K9me3 at the *S. pombe* insert region as visualised by the FISH signal (green channel). DNA is stained with DAPI (blue channel). Scale bar in **A** indicates 5 µm. The site of *S. pombe* DNA is indicated in **B** by yellow arrows.
**C** - Average chromosome profile for H3K9me3 on the NP-A4 insert-bearing chromosome across several images (n=31). Immunolocalisation (red) and FISH (green) signal intensity were measured along the length of the chromosomes and binned according to their position before plotting. Error bars represent +/- SEM. Average H3K9me3 stain intensity was compared between the regions of 45-55% (endogenous mouse DNA) and 75-85% (*S. pombe* DNA corresponding to the highest FISH signal) by the KS test (*p<0.01).

results in suggesting that the *S. pombe* chromatin carried by the NP-A4 cell line is heterochromatic. Together, these results show that in both mouse and human cells heterochromatin readily assembles over inserted *S. pombe* chromatin within 30 cell divisions

## 5.6 Discussion

To further characterise the *S. pombe* chromatin inserted in F1.1 the levels and distribution of various chromatin marks across the insert were analysed. By both ChIP and immunolocalisation the *S. pombe* chromatin insert in F1.1 was observed to have high levels of H3K9me3, the distinguishing mark of constitutive heterochromatin (**Figure 5.1**). Accordingly, high levels of HP1α and 5meC and low levels of activating marks were also found across the *S. pombe* region (**Figures 5.1-5.3**). The sheer size of the *S. pombe* insert across which the heterochromatin has formed makes this quite a distinctive stretch of chromatin.

The main question that arises from these analyses is what triggers heterochromatin formation over the inserted *S. pombe* DNA. The most likely explanation is that some feature of the *S. pombe* chromatin leads to establishment of H3K9me3 heterochromatin after its insertion. These features could lie within the DNA itself, or within the chromatin proteins inserted with it. A few possibilities therefore present themselves:

1. DNA features:

a. Certain sequence features of the *S. pombe* DNA could be leading to the recruitment of chromatin modifiers or heterochromatin proteins to the *S. pombe* insert, although it is unclear what these sequence features might be.

b. The unusual DNA methylation profile of *S. pombe* DNA could target it for heterochromatinisation. *S. pombe* has no detectable DNA methylation, while the majority of the genome is CpG methylated in mouse (Capuano *et al.* 2014, Bird 2002). The introduction of a large stretch of unmethylated DNA into the mouse cell could thus serve as a recognition signal for foreign chromatin, or simply as a blank platform for certain DNA binders such as chromatin modifiers, leading to the

establishment of heterochromatin. The F1.1 *S. pombe* DNA was observed to become gradually methylated after its insertion into the mouse genome (McManus *et al.* 1994), which may be an indicator of the gradual establishment of heterochromatin over the insert.

2. Chromatin features:

c. Heterochromatin spreads across the *S. pombe* insert from pre-existing domains of heterochromatin in the *S. pombe* cell (**Figure 5.8**). The spreading of heterochromatin to adjacent regions in *cis* is thought to be a primary mechanism for its propagation in many species (Allshire and Madhani 2017). Whole genome sequencing of F1.1 shows that several regions of DNA associated with constitutive H3K9me3 heterochromatin in *S. pombe* were inserted in F1.1, including centromeric and telomeric regions. Given that the F1.1 cell line was created by direct fusion of C127 with *S. pombe* protoplasts, these regions would have been introduced to the mouse cell already assembled on heterochromatin. It is possible then that these heterochromatin domains served as a platform within the *S. pombe* insert from which heterochromatin spread to adjacent regions, eventually covering the entirety of the *S. pombe* chromatin. The spread of heterochromatin is normally counteracted by the presence of barriers to its transmission, which include regions of transcription, activating marks and chromatin barriers such as insulator proteins. However, rearrangements of the DNA sequence have been known to lead to heterochromatin spreading onto normally euchromatic loci from newly adjacent heterochromatic loci, such as has been well documented in *Drosophila* position effect variegation (PEV) (Elgin and Reuter 2013) or, in the case of facultative heterochromatin, for mammalian X:autosome translocations (Sharp *et al.* 2002). Furthermore, the high level of divergence between the *S. pombe* and mammalian genomes means that any barriers to heterochromatin spreading in the *S. pombe* genome are unlikely to be recognised in a mammalian cell. Without the presence of such factors as recognisable promoters, enhancers or CTCF sites, it is likely that the heterochromatin would spread over the entire *S. pombe* chromatin domain in the mammalian cell without much hindrance.

The fusion method by which the mammalian-*S. pombe* hybrid cells were generated does not allow us to distinguish between the influence of DNA sequence or

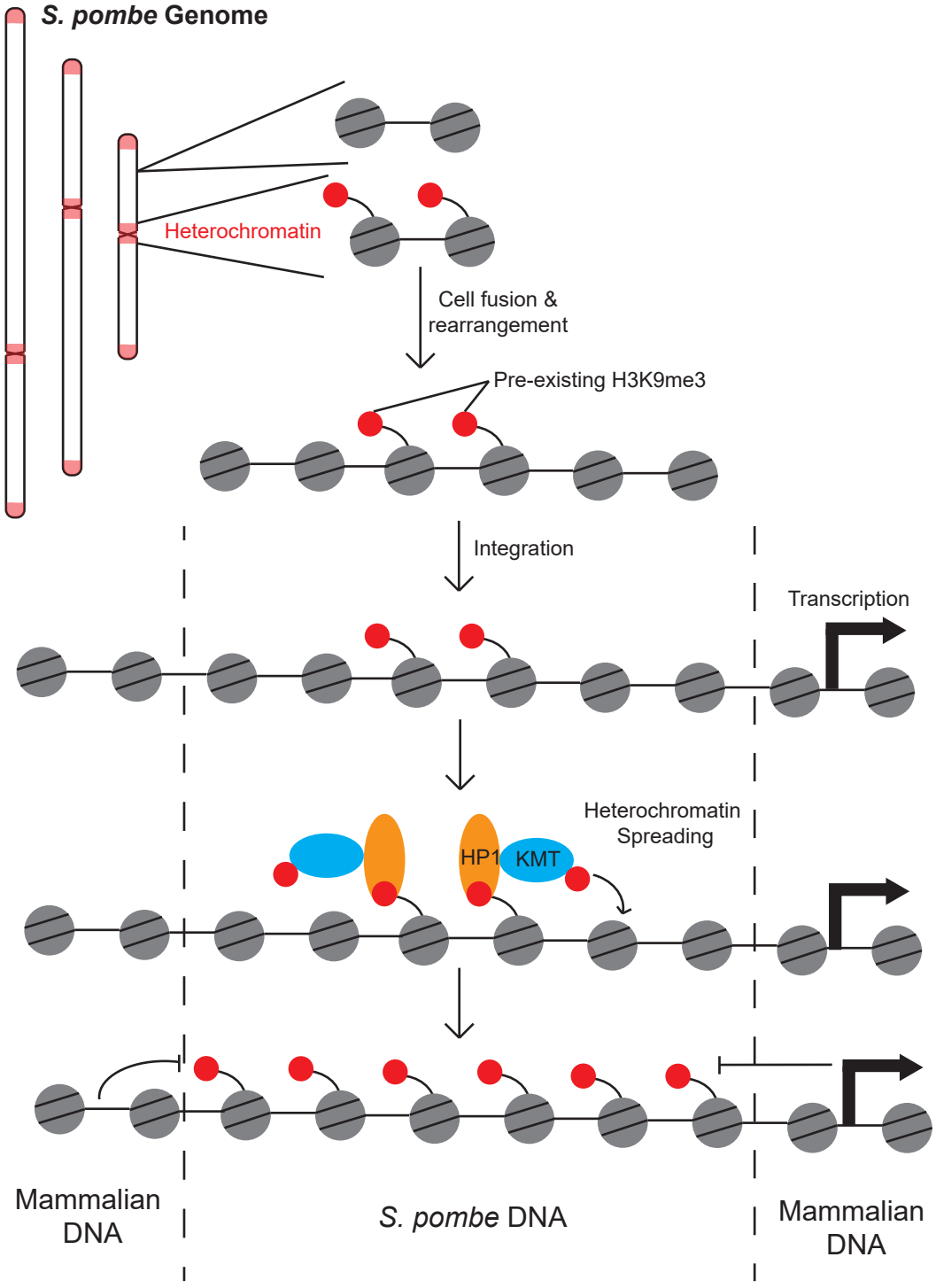


**Figure 5.8 - Heterochromatin spreads from pre-existing *S. pombe* heterochromatin in mammalian fusion cells**

Model for the establishment of heterochromatin over *S. pombe* DNA in mammalian cells. *S. pombe* DNA is inserted as chromatin by cell fusion and certain regions maintain their heterochromatic identity from the *S. pombe* cell. This heterochromatin is recognised and propagated by the mammalian proteins to establish a heterochromatic domain over the *S. pombe* DNA. Heterochromatin spreading is halted at mammalian DNA by barrier signals not present within the *S. pombe* DNA which include regions of active transcription.

modifications and chromatin proteins. A new experimental set-up that allows uncoupling of the effects of DNA features on the one hand, and chromatin proteins on the other, is therefore highly desirable.

The heterochromatin spreading hypothesis opens up the question of where along the chromosome the spreading is halted. In PEV the spreading of heterochromatin is usually confined to one or a few genes, while X:autosome translocations have been shown to have a variable effect on the translocated autosome, with large regions of chromatin sometimes “escaping” from silencing (Elgin and Reuter 2013, Duthie *et al.* 1999, Disteche 1999, Dobigny *et al.* 2004). Is heterochromatin confined to the region of *S. pombe* chromatin? If so, to what extent is the insert region covered by heterochromatin? By light microscopy on metaphase spreads the H3K9me3 and FISH signals appeared to be tightly correlated, suggesting that the heterochromatin is only associated with *S. pombe* DNA (**Figure 5.1**). ChIP of the mouse sequences flanking the *S. pombe* insert site supported this view, showing no encroachment of heterochromatin into the surrounding chromatin (**Figure 5.4**). Given that the order of the *S. pombe* DNA sequences within the insert remains unknown, it was not possible to correlate the levels of H3K9me3 on a particular region of *S. pombe* DNA with that region’s position within the insert. It is possible that the H3K9me3 levels remain high on the *S. pombe* chromatin right up to the junction with the mouse chromatin. Alternatively, the H3K9me3 levels might gradually decrease as the sequence approaches the junction such that the *S. pombe* sequences in close proximity to the mouse sequence are no longer heterochromatic. While we are not able to distinguish between these two possibilities, both ChIP and immunolocalisation suggest that the H3K9me3 is widespread within the *S. pombe* insert, so it can be said with certainty that the majority of the *S. pombe* chromatin in F1.1 is heterochromatic, while the surrounding mouse chromatin is not.

One region in which the *S. pombe* chromatin is less heterochromatic is around the SV2NEO gene. Expression of the gene is under constant selection by G418 treatment, and low-level transcription from the gene is detectable in F1.1 cells (**Figure 5.5**). Levels of activating histone marks show a small increase at the SV2NEO gene, suggesting there may be an island of permissive chromatin around the gene within the larger heterochromatic domain (**Figure 5.3**). Low levels of

transcription are also detectable at the nearby *ura4*$^{+}$ locus into which the SV2NEO gene was inserted in the original *S. pombe* strain. However, levels of activating histone marks at the *ura4*$^{+}$ locus are low and indistinguishable from the rest of the *S. pombe* chromatin, suggesting that if there is an island of permissive chromatin around SV2NEO it is very localised. Furthermore, H3K9me3 and HP1 levels at SV2NEO remain very high (**Figure 5.1**). It therefore appears as though the chromatin around the SV2NEO gene has struck a balance between the heterochromatin seeking to spread from nearby regions and the active chromatin seeking to establish itself to allow for transcription. This would explain why such low levels of expression are observed from what is ostensibly a strong promoter driving SV2NEO expression.

The NP-A4, HeP2 and HeP3 cell lines, made from *S. pombe* fusion with either mouse NIH-3T3 or human HeLa cells, also display similar high and widespread levels of H3K9me3 across their respective *S. pombe* inserts (**Figures 5.6 & 5.7**). The *S. pombe* chromatin of both the NP-A4 and HeP3 cell lines has therefore been shown to share both the distinctive structural appearance and the heterochromatic signature of the F1.1 *S. pombe* chromatin. These two chromatin features therefore appear to correlate in the three mammalian-*S. pombe* fusion cells examined thus far. An important question will therefore be to determine if this link is meaningful and if, for instance, the heterochromatic signature somehow influences the mitotic chromosome structure.

# Chapter 6: *S. pombe* DNA inserted directly into mammalian cells does not assemble heterochromatin or adopt a distinct appearance

## 6.1 Introduction

In the previous chapter the possibility of a link between the H3K9me3 mark and mitotic chromosome structure was established. These two properties of the chromatin were found to be correlated in all three mammalian – *S. pombe* fusion cells created: F1.1, NP-A4 and HeP-3. These results suggest the hypothesis that H3K9me3 is the cause of the distinct chromatin structure observed at *S. pombe* chromatin inserted in mammalian cells. Important evidence that would further strengthen this hypothesis would be to find the reverse of what has already been observed, i.e. the correlation of the absence of these two properties. The two main ways to test this hypothesis would be as follows:

1. The best test for a link between H3K9me3 and the distinct chromatin structure of the *S. pombe* chromatin would involve the removal of the proposed upstream property (H3K9me3) to observe the effect it may have on the downstream property (chromatin structure). If H3K9me3 were removed from the *S. pombe* chromatin in F1.1, NP-A4 or HeP-3 and the mitotic structure of the region were observed to revert to an appearance similar to that of mouse chromatin, this would be strong evidence for a causative link between the two properties. If, on the other hand, removal of H3K9me3 did not result in a change in structure, then such a link would be disproved. This test therefore has the benefit of providing a conclusive result in either case, dependent upon the successful removal of H3K9me3 from the *S. pombe* chromatin. However, implementing a system that specifically removes H3K9me3 from the *S. pombe* insert could prove challenging.

2. A second test would be the generation of new mammalian cell lines containing *S. pombe* chromatin that either does not assemble the distinct structure, does not possess high levels of H3K9me3, or both. Thus far, all the cell lines observed have possessed both H3K9me3 and a distinct structure over their respective *S. pombe* inserts. If a region of *S. pombe* chromatin were observed to display one of these two

properties but not the other, then the two properties will have been separated, thereby disproving a causative link. If on the other hand a region of *S. pombe* chromatin showing neither H3K9me3 nor a distinct structure were observed, this would provide additional evidence for a correlation between these properties. While not as conclusive as the removal of H3K9me3 from one of the existing cell lines, this would nonetheless provide an example of the correlation of the absence of these properties to reinforce the correlation of their presence.

Both of these approaches would also rule out any direct influence of the DNA sequence itself on the distinct chromatin structure of the *S. pombe* insert. In the first test the sequence would be exactly the same before and after H3K9me3 removal. In the second test the source of the DNA would be the same as in the other cell lines F1.1, NP-A4 and HeP3. The experiments described in this chapter therefore set out to achieve one or both of these two goals: the selective removal of H3K9me3 from one of the existing cell lines, and the generation of a new cell line without H3K9me3 and/or a distinct structure over *S. pombe* chromatin.

## 6.2 F1.1 cells are not susceptible to reprogramming by standard iPS methods or cell fusion

In a first attempt to remove H3K9me3 from the *S. pombe* insert chromatin, F1.1 cells were subjected to standard methods of reprogramming into induced pluripotent stem cells (iPS). Since the first demonstration that somatic cells could be reprogrammed into a pluripotent state by the expression of four key transcription factors, the technique of iPS reprogramming has become an important method in the study of stem cells and their therapeutic potential (Takahashi and Yamanaka 2006, Takahashi *et al.* 2007, Okita *et al.* 2007). Embryonic stem cells (ESCs) have long been known to have an overall more open and euchromatic profile that somatic cells based on such observations as the increase in heterochromatin levels over the course of differentiation (Meshorer *et al.* 2006, Efroni *et al.* 2008, Becker *et al.* 2016). iPS cells have been found to be highly similar to ESCs in terms of their chromatin marks, and studies of the reprogramming process have shown clear decreases in H3K9me3 and DNA methylation and increases in activating chromatin marks during reprogramming (Bock *et al.* 2011, Mikkelsen *et al.* 2008, Mattout *et al.*

2011, Koche *et al.* 2011). Reprogramming of the F1.1 cell line might therefore lead to a reduction of heterochromatin over the *S. pombe* chromatin.

To reprogram the F1.1 cells a virus-free technique was employed whereby the four reprogramming factors Oct4, Sox2, cMyc and Klf4 are expressed from a plasmid which is stably inserted into the target genome by a *PiggyBac* transposon (Yusa *et al.* 2009, Kaji *et al.* 2009). Successfully transfected cells were identified by expression of a GFP marker gene and reprogrammed cells can be easily identified by their ESC-like morphology and colony formation. The construct was successfully inserted into F1.1 and, after a few days of growth in ESC media, some GFP-positive cells began to exhibit what appeared to be partial reprogramming, with a more rounded morphology than normal F1.1 cells (**Figure 6.1 A**). However, after several more days growth these cells did not reprogram fully and even reverted back to their usual somatic morphology. Reprogramming of both F1.1 and the parental C127 cell line was also attempted by introducing the reprogramming factors via lentivirus. Once again however, no cells with stem cell morphology were observed with either cell line after two weeks of growth in ESC media.

The distinctly different approach of reprogramming by fusion with ESCs was also attempted. Fusion of somatic and pluripotent cells has been shown to reprogram the somatic nucleus in a similar way to iPS methods, resulting in stem cell-like fusion products (Cowan *et al.* 2005, Han *et al.* 2011, Foshay *et al.* 2012). This has been attributed to the presence of the reprogramming factors in stem cells which act on the somatic nucleus after fusion. To test this method, F1.1 cells were fused with mouse embryonic E14 stem cells containing a puromycin resistance selectable marker. Fusion products were isolated by selection in both G418 and puromycin in stem cell media. A few doubly-resistant fusion products were isolated, but none of them progressed to adopt a stem cell morphology. RNA from two such products was extracted and expression levels of a few key stem cell markers was determined by RT-qPCR and compared to F1.1 or E14 expression levels (**Figure 6.1 B**). Both hybrids showed an expression profile comparable to F1.1 and not at all similar to E14 stem cells. Notably, the hybrid cells completely lacked expression of the key stem cell genes Eras and Nanog, which were highly expressed in E14. The fusion products thus maintained their F1.1 / C127 somatic identity and failed to adopt a stem cell profile.

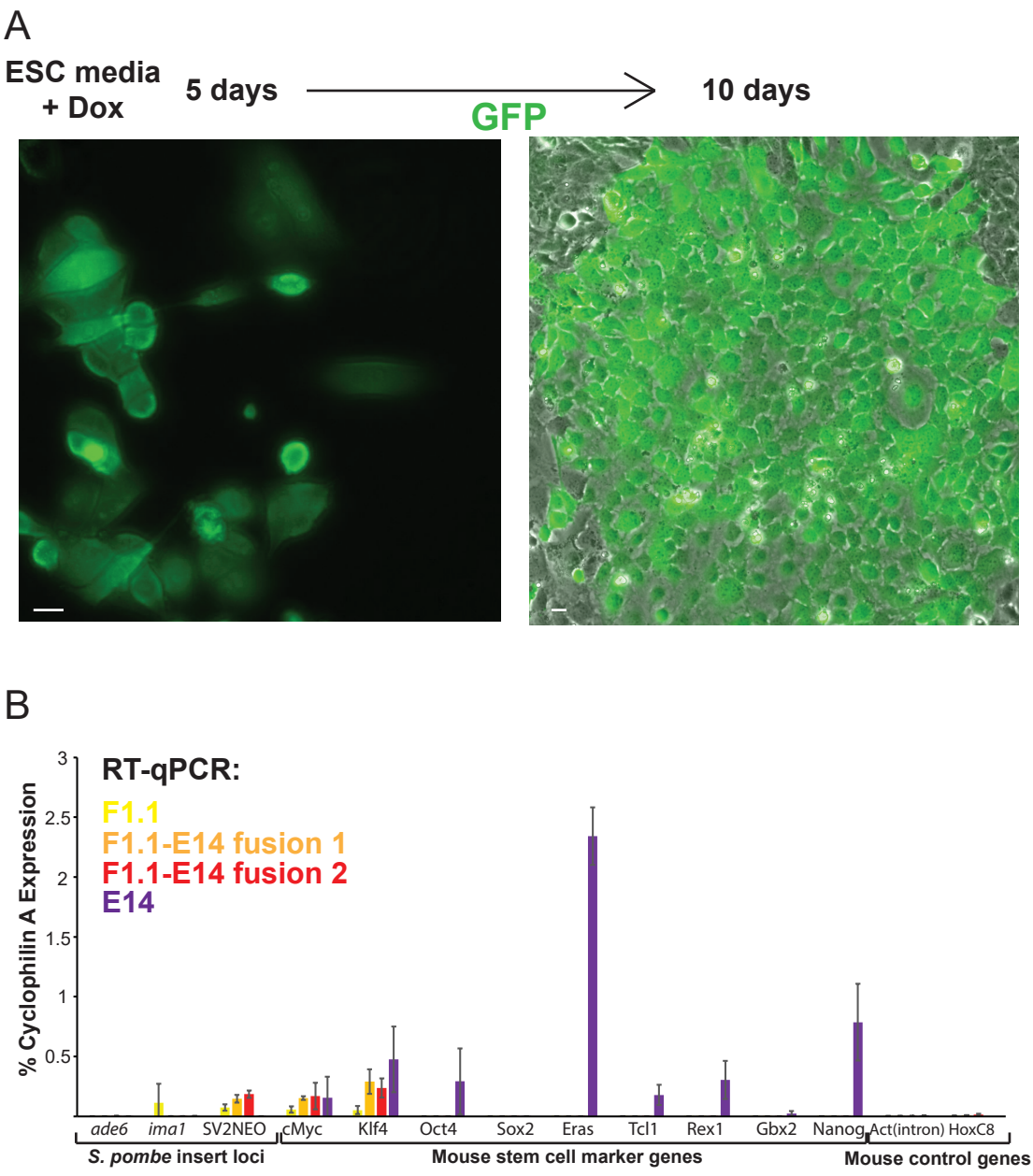


**Figure 6.1 - The F1.1 cell line is not susceptible to reprogramming by iPS induction or cell fusion**

**A** - Microscopy images showing F1.1 cells transfected with plasmids expressing Oct4, Sox2, cMyc and Klf4 alongside a GFP reporter gene. GFP positive cells should reprogram to iPS cells over time in ESC media with doxycycline. After 5 days in dox GFP positive cells begin to adopt stem cell-like morphology. However after an additional 5 days GFP positive cells have reverted back to a somatic appearance and have not formed colonies. Scale bars indicate 15 μm.
**B** - RT-qPCR of total RNA from F1.1 and the E14 embryonic stem cell line as well as two independently derived cell lines resulting from fusion of F1.1 and E14. Expression levels of certain stem cell markers indicate stem cell identity. An intronic region of mouse actin and the lowly expressed HoxC8 gene are included as controls. Expression levels were normalised to the highly expressed mouse Cyclophilin A gene. Error bars represent +/- SEM of three independent repeats.

Both iPS methods and cell fusion were thus unsuccessful in reprogramming F1.1 / C127 cells to a pluripotent state. An alternative approach for the removal of H3K9me3 heterochromatin from the *S. pombe* chromatin was therefore necessary.

## 6.3 Partial removal of H3K9me3 by targeted demethylation does not alter the structure of the *S. pombe* insert

To more directly investigate the potential role of heterochromatin in forming the unusual chromatin structure over the *S. pombe* insert, the H3K9me3 mark was selectively targeted for removal in F1.1 cells. For this, the synthetic PREditOR system for global H3K9me3 removal in cells was used (Molina *et al.* 2016). Expression of a construct consisting of the H3K9me-binding chromodomain of Suv39H1 fused with the H3K9-demethylase jmjC domain of JMJD2D and EYFP was shown to effectively remove H3K9me3 from centromeric regions of HeLa cells. This construct could therefore be employed as a tool to remove H3K9me3 from the *S. pombe* insert in F1.1 cells.

To allow for tight control of expression in F1.1 cells, the Suv39CD-EYFP-JMJD2D construct was stably inserted into the F1.1 genome using a Sleeping Beauty transposon, with a Tet-inducible promoter driving its expression. Control cell lines expressing Suv39CD-EYFP and Suv39CD*-EYFP-JMJD2D constructs that lack the demethylase activity and the targeting ability, respectively, were also generated (**Figure 6.2**). Expression of the constructs was successfully induced upon treatment with the tetracycline analogue doxycycline (dox) (**Figure 6.3**). The localisation of the constructs was as expected for each: Suv39CD-EYFP and Suv39CD-EYFP-JMJD2D were concentrated in regions of high H3K9me3 while Suv39CD*-EYFP-JMJD2D, which cannot bind H3K9me3, was distributed throughout the nucleus. The effect of each construct on H3K9me3 levels was also as expected. After 48h of expression global H3K9me3 levels appeared to be reduced in cells expressing Suv39CD-EYFP-JMJD2D, and to an extent Suv39CD*-EYFP-JMJD2D, but not Suv39CD-EYFP (**Figure 6.3**).

To determine if H3K9me3 were reduced not only globally but specifically over the region of *S. pombe* insertion, ChIP-qPCR was performed on the three resulting cell

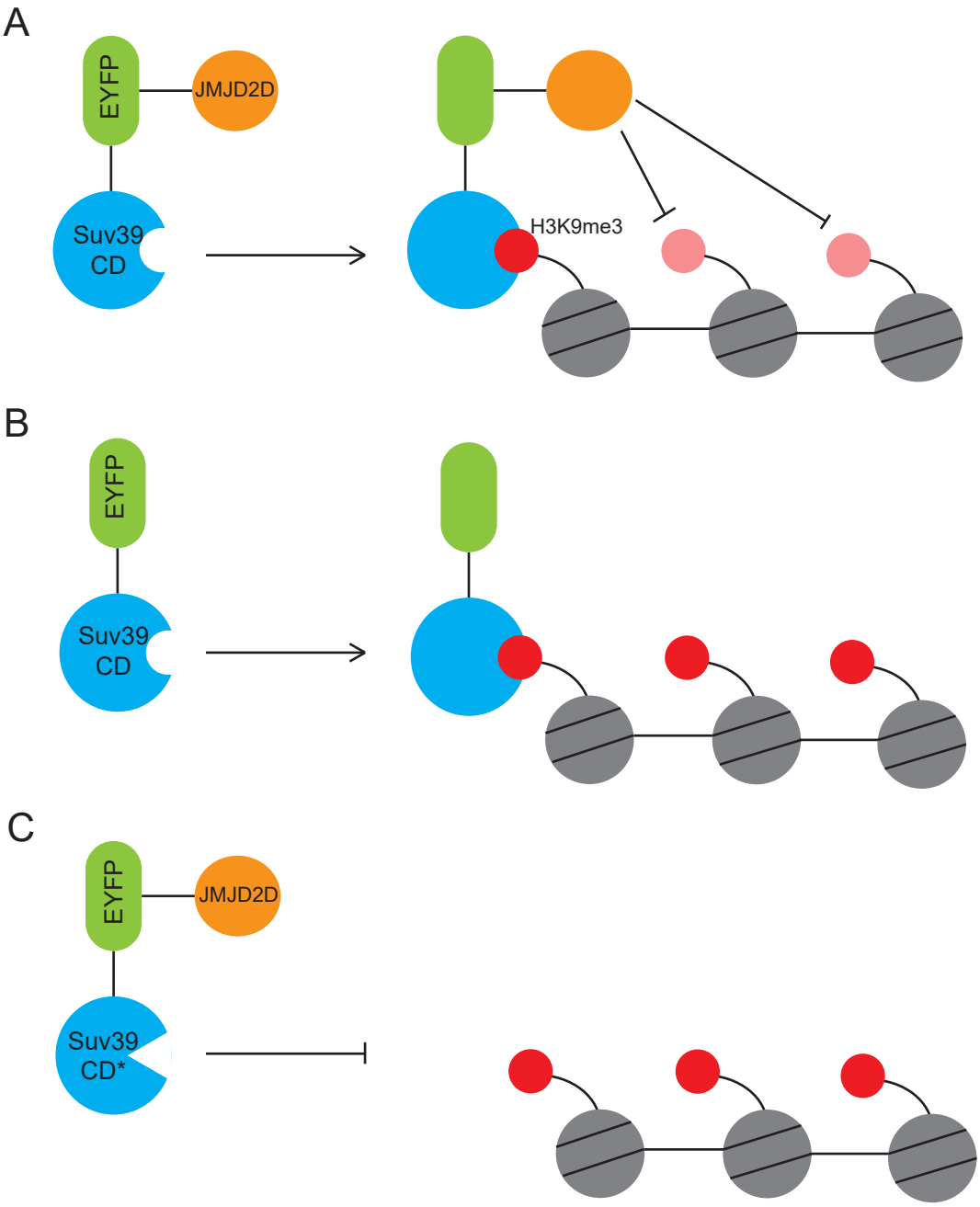


**Figure 6.2 - Removal of H3K9me3 by the PREditOR system**

Diagram of the action of the synthetic PREditOR constructs.
**A** - Suv39CD-EYFP-JMJD2D binds to and removes H3K9me3
**B** - Suv39CD-EYFP binds to H3K9me3 but has no demethylase activity to remove it
**C** - Suv39CD*-EYFP-JMJD2D has demethylase activity but no ability to bind H3K9me3.

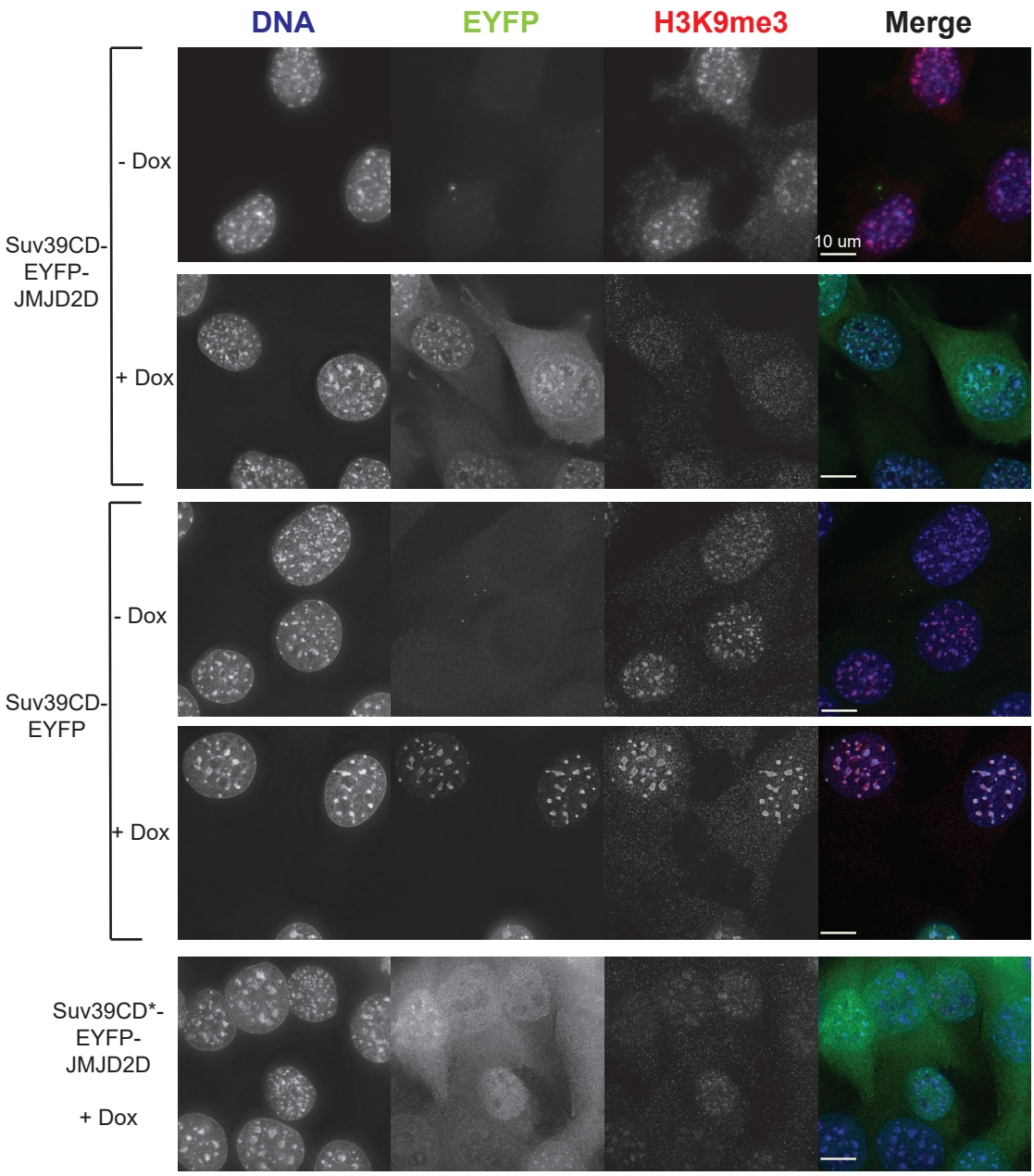


**Figure 6.3 - Induction of Suv39CD-EYFP-JMJD2D expression in F1.1 results in decreased H3K9me3 levels**

Microscopy images of F1.1 interphase cells with the indicated PREDitOR construct stably integrated: Suv39CD-EYFP-JMJD2D (fully active), Suv39CD-EYFP (no demethylase activity) or Suv39CD*-EYFP-JMJD2D (no targeting activity). Cells were grown in either regular media (- Dox) or with doxycycline for 48h (+ Dox). DNA is stained with DAPI (blue channel) and immunolocalisation for H3K9me3 was performed on cells (red channel). PREditOR expression is indicated by EYFP (green channel). Scale bars indicate 10 μm.

lines after 72h of dox treatment (**Figure 6.4 A**). These results confirm the ability of the construct to remove H3K9me3 from pericentromeric regions, with F1.1 cells expressing Suv39CD-EYFP-JMJD2D showing a more than 6-fold decrease in H3K9me3 levels at major satellites compared with cells expressing the control Suv39CD-EYFP construct. However, removal of H3K9me3 from regions within the *S. pombe* insert was not as effective. H3K9me3 levels at *S. pombe* loci were only reduced by half in Suv39CD-EYFP-JMJD2D cells relative to Suv39CD-EYFP cells, a decrease which was not significant across the three replicates. The untargeted control Suv39CD*-EYFP-JMJD2D showed a pattern intermediate between the other two constructs. This likely reflects the fact that the demethylase activity of this construct remains intact, and its overexpression could lead to reduced H3K9me3 throughout the genome. However, the lack of targeting via its CD domain prevents the H3K9 demethylation from being as efficient as for the targeted Suv39CD-EYFP-JMJD2D construct. These results suggest the *S. pombe* insert is less permissive than centromeres to H3K9me3 removal by Suv39CD-EYFP-JMJD2D.

These results were confirmed by microscopy (**Figure 6.4 B**). H3K9me3 levels were assessed by immunolocalisation, measuring the intensity of the fluorescent signal as before, and correlated with FISH for *S. pombe* DNA. Control cells expressing Suv39CD-EYFP or Suv39CD*-EYFP-JMJD2D show the normal profile of H3K9me3 along the insert-bearing chromosome 10, with high levels of H3K9me3 at both the centromere and over the region containing *S. pombe* DNA. In cells expressing the fully active Suv39CD-EYFP-JMJD2D however, H3K9me3 levels at the centromere are strongly reduced. Indeed, H3K9me3 fluorescence intensity was more or less constant from the centromere to the middle of the chromosome arm, indicating that H3K9me3 levels at centromeres were reduced to background levels. However, high H3K9me3 remained over the region containing *S. pombe* DNA. While the peak of H3K9me3 over the *S. pombe* DNA was not as high compared to background or as sharp as in control cells, it remained statistically significant compared to the surrounding chromatin. Thus, both microscopy and ChIP analysis show that H3K9me3 is effectively removed by Suv39CD-EYFP-JMJD2D at centromeres, but only partially reduced over the region containing *S. pombe* DNA.

This incomplete removal of H3K9me3 over the *S. pombe* insert makes it difficult to draw conclusions as to the effect of H3K9me3 on the insert structure. Examination

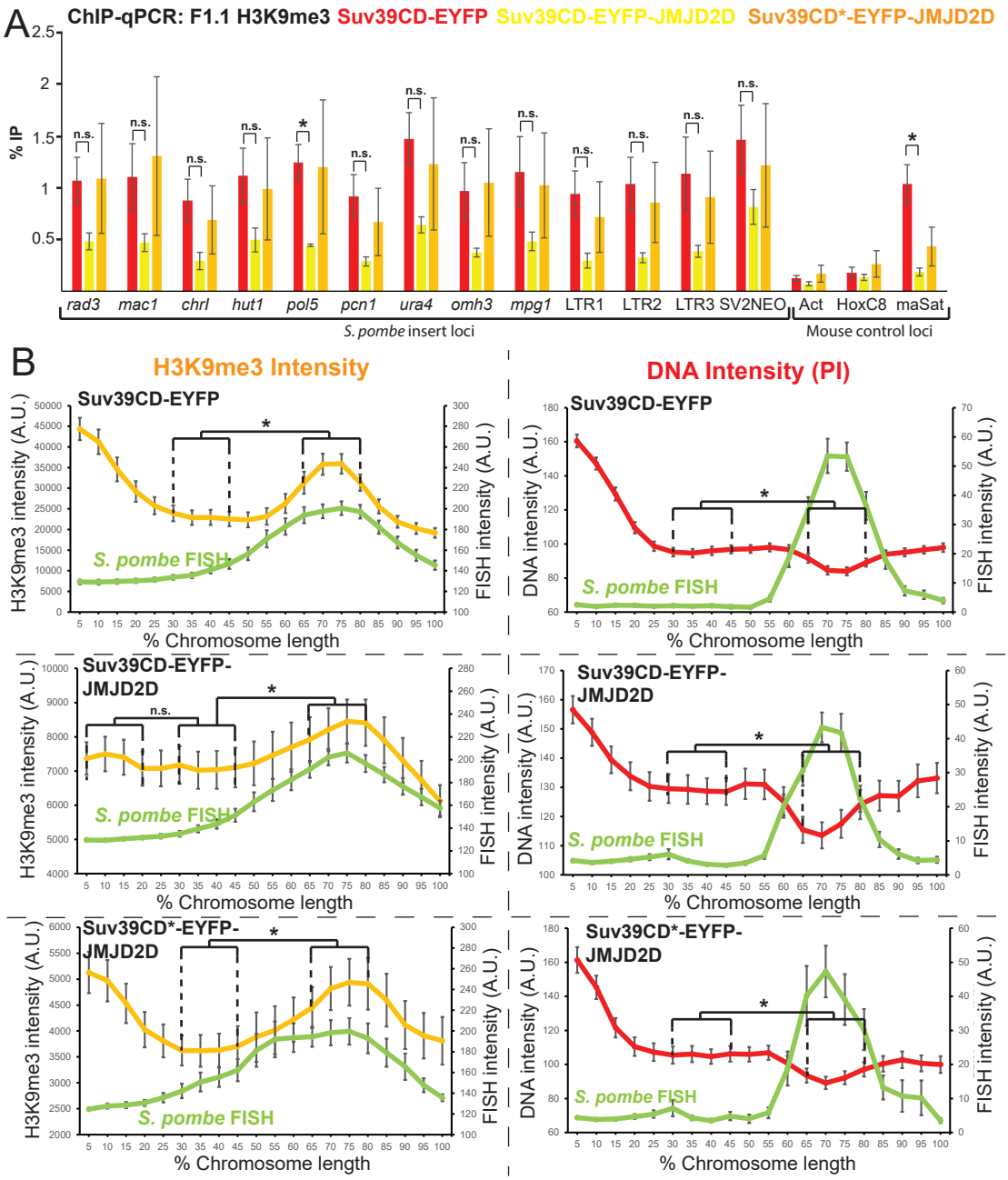


**Figure 6.4 - Partial removal of H3K9me3 from the F1.1 *S. pombe* DNA insert has no visible effect on its structure**

**A** - ChIP-qPCR for H3K9me3 at 13 loci within the *S. pombe* insert and 3 mouse control loci in F1.1 interphase cells expressing one of the PREditOR constructs for 72h. Suv39CD-EYFP-JMJD2D targets the H3K9me3 for removal, Suv39CD-EYFP lacks H3K9me3 demethylase activity and Suv39CD*-EYFP-JMJD2 lacks H3K9me3 targeting ability. Act is the highly transcribed negative control gene actin, HoxC8 is a region of facultative heterochromatin and maSat is the satellite positive control region. Error bars represent +/- SEM of three independent repeats. %IP levels were compared by a *t*-test (*p<0.05, n.s. = not significant). **B** - Chromosome profile for H3K9me3 instensity (orange, left) or DNA intensity (red, right) measured by microscopy on the F1.1 insert bearing chromosome, with insert location shown by FISH intensity (green). Cells were expressing the indicated PREditOR constructs for 72h. From top to bottom n = 31, 25 and 17 for H3K9me3 and n = 59, 45 and 26 for DNA. Error bars represent +/- SEM. Average H3K9me3 and PI intensities were compared between the regions of 30-45% (mouse DNA) and 65-80% (*S. pombe* DNA) by the KS test (*p<0.01).

of the DNA stain intensity over the insert-bearing chromosome 10 in cell lines expressing any of the three different constructs shows no change in *S. pombe* insert chromatin structure (**Figure 6.4 B**). All three cell lines show the same distinctive dip in DNA stain intensity coinciding with the region of *S. pombe* FISH. However, it remains unclear if the 50% decrease in H3K9me3 levels at the *S. pombe* insert which we observe would be sufficient to reverse the structural changes in mitosis. In order to confidently determine if H3K9me3 does or does not alter mitotic chromosome structure a much more robust reduction in H3K9me3 levels would be required.

## 6.4 Naked *S. pombe* DNA inserted into mammalian cells does not assemble heterochromatin or display a distinct structure

As discussed **Chapter 5**, the *S. pombe* DNA inserted into the F1.1, NP-A4 and HeP3 cell lines was introduced as fully formed chromatin by cell fusion. Two possibilities for the origin of the heterochromatin and distinct structure present themselves:

1. *S. pombe* DNA alone induces the formation of heterochromatin and/or the distinct mitotic structure observed in these cells

2. Specific aspects of *S. pombe* chromatin, not DNA, lead to formation of heterochromatin and a distinct structure.

To distinguish between these possibilities, large segments of *S. pombe* DNA devoid of chromatin proteins were introduced by transfection to NIH-3T3 cells to generate mouse cell lines with stably integrated *S. pombe* DNA. G418 resistant cells were screened by FISH for large regions of *S. pombe* DNA inserted at a single locus within a mouse chromosome, resulting in the isolation of several such cell lines (**Figure 6.5**).

In most of these DNA transfected cell lines the *S. pombe* DNA inserted adjacent to a centromere or telomere, making analysis of the structure by DNA stain intensity difficult to assess. However, one cell line, NP-D4, possessed a large region of *S. pombe* DNA inserted about halfway along a chromosome arm. Not only was the *S.*

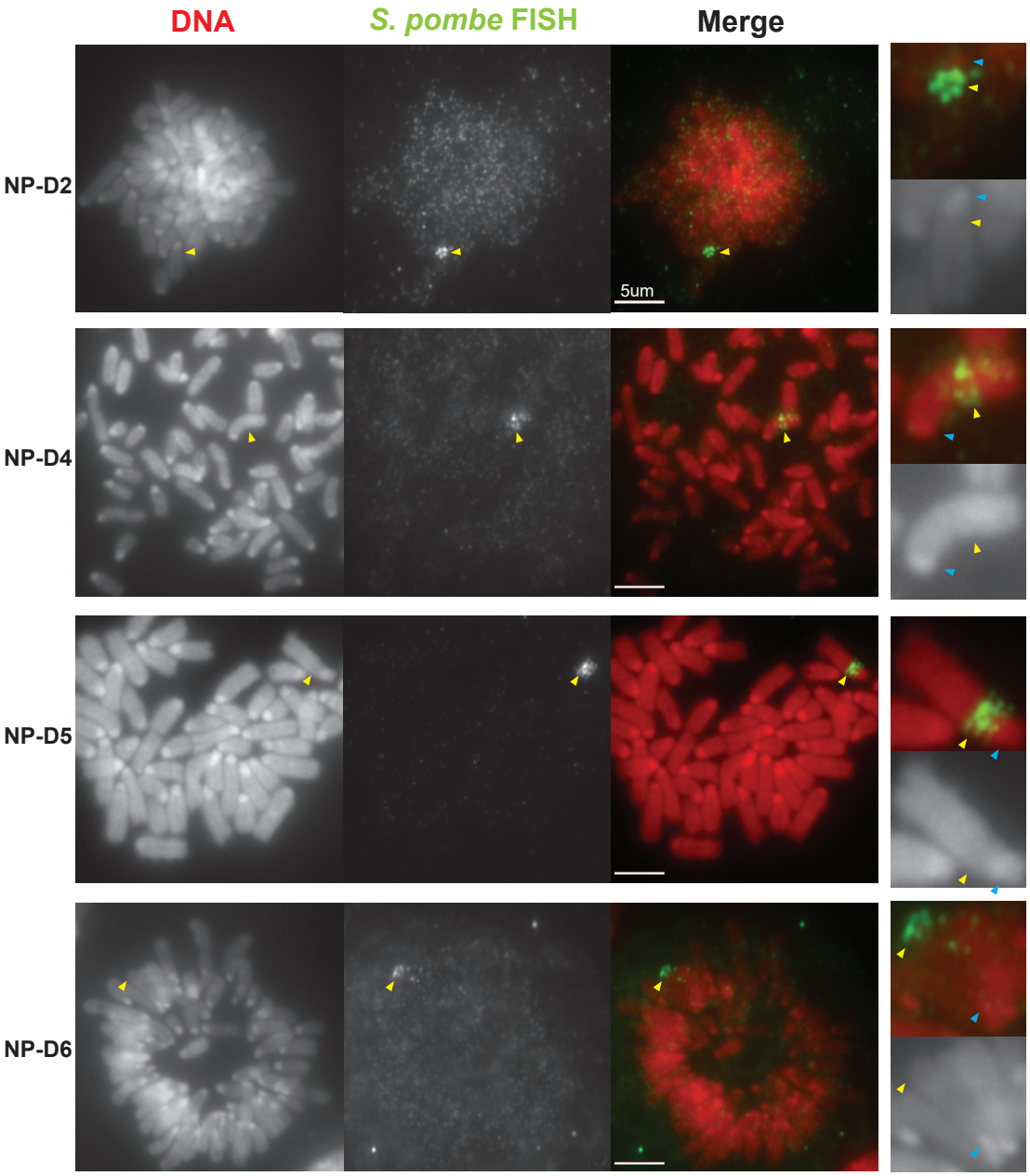


**Figure 6.5 - Large insertions of *S. pombe* DNA in mouse chromosomes by transfection**

Metaphase spreads of four NIH-3T3 derived cell lines containing stable integrations of transfected S. pombe DNA. Detailed images of insert bearing chromosomes are shown to the right of the larger images. DNA is stained with PI (red channel) and *S. pombe* DNA is highlighted by FISH using biotinylated total *S. pombe* DNA bound to avidin-FITC (green channel) and indicated by yellow arrows. Centromeres appear as regions of bright DNA staining and are indicated by blue arrows. Scale bars indicate 5μm.

*pombe* region in NP-D4 similar to the F1.1 insert in terms of its relative position on a similarly sized mouse chromosome, the insert appeared to be of a similar size to that seen in F1.1. Thus, NP-D4 was the best candidate among the DNA transfectants for comparison of mitotic chromatin structure with that of F1.1.

In contrast to the insert-bearing chromosomes of F1.1, NP-A4 and HeP-3, analysis of the DNA stain intensity over the *S. pombe* DNA insert-bearing chromosome in NP-D4 showed no variation across the entire chromosome, with the exception of the expected peak at the centromere (**Figure 6.6**). The *S. pombe* DNA in the NP-D4 DNA transfectant therefore does not appear to have adopted a distinctive chromatin structure, and appears indistinguishable from the surrounding endogenous chromatin by the same approach that identified the unusual structure in the fusion derived cell lines. This suggests that the difference may be a consequence of the DNA transfection method used to generate NP-D4, as opposed to the protoplast fusion method that generated F1.1, NP-A4 and HeP-3.

To determine if H3K9me3 is concentrated over the *S. pombe* DNA in NP-D4, the levels of H3K9me3 were measured by ChIP-qPCR. H3K9me3 levels were around 3-fold lower on *S. pombe* regions analysed in NP-D4 compared to the fusion cell lines F1.1 and NP-A4. (**Figures 5.1, 5.6**, **6.7 A**). This represented a significantly lower H3K9me3 level across the *S. pombe* insert compared to pericentromeric satellite DNA in NP-D4. The lack of heterochromatin formation on the *S. pombe* region in NP-D4 was even more evident by microscopy. Immunolocalisation for H3K9me3 on NP-D4 metaphase spreads showed no increase in H3K9me3 over the *S. pombe* insert compared to surrounding endogenous chromatin (**Figure 6.7 B-D**). This is in contrast to F1.1 which displayed a clear region of high H3K9me3 coincident with the *S. pombe* DNA. Unlike all other cell lines examined, the *S. pombe* insert in NP-D4 is indistinguishable from surrounding chromatin both in terms of its mitotic structure and its H3K9me3 levels. While the F1.1, NP-A4 and HeP-3 cell lines showed a correlation between the presence high levels of H3K9me3 and a distinct chromatin structure, the NP-D4 cell line therefore suggests that mitotic chromosome structure is not altered in the absence of H3K9me3 heterochromatin assembled on the *S. pombe* DNA insert.

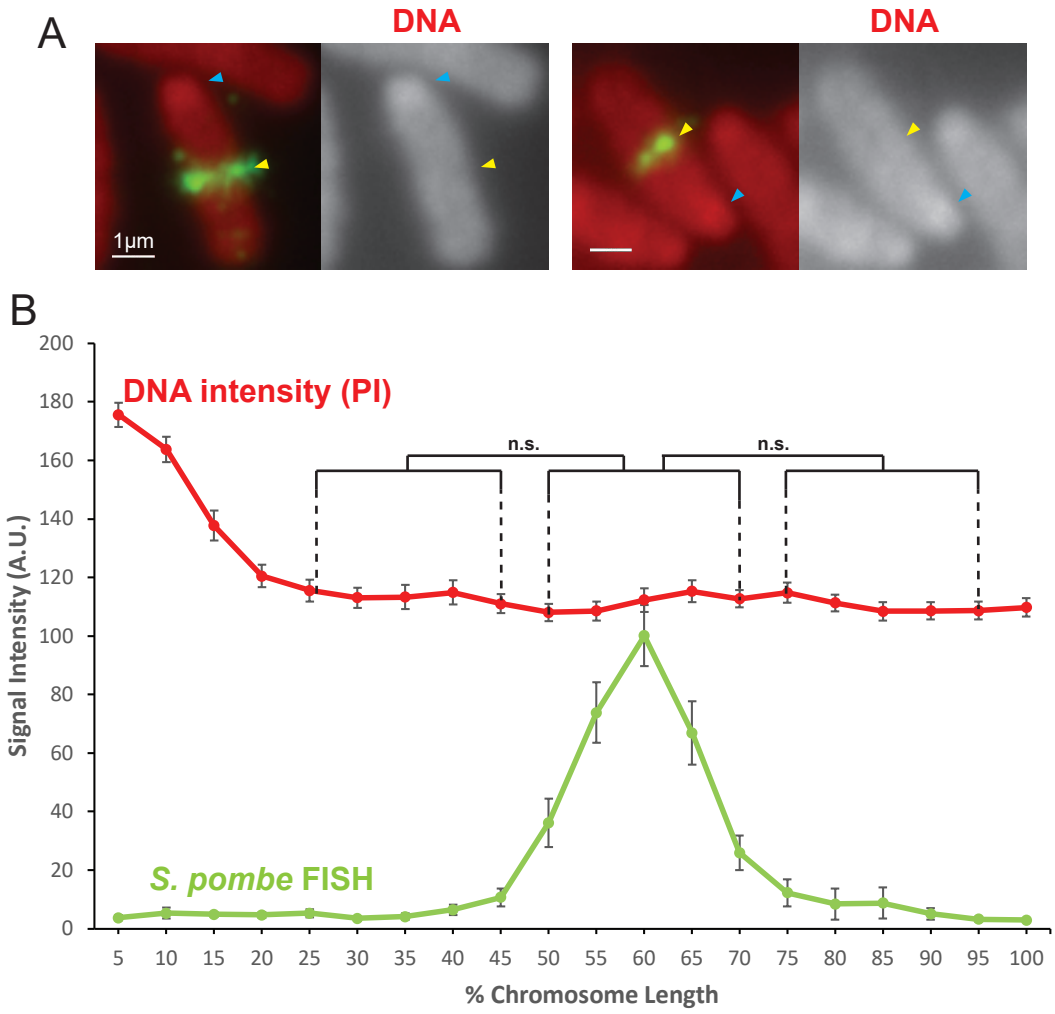


**Figure 6.6 - No distinct metaphase chromosome structure is detectable over the *S. pombe* DNA in NP-D4**

**A** - Metaphase spreads showing insert bearing chromosomes of the transfection-derived NP-D4 cell line. DNA is stained with propidium iodide (PI) and *S. pombe* DNA is highlighted by FISH using biotinylated total *S. pombe* DNA bound to avidin-FITC and indicated by yellow arrows. Centromeres appear as regions of bright DNA staining and are indicated by blue arrows. Scale bars indicate 1µm.
**B** - Average chromosome profile of the NP-D4 insert bearing chromosome across several images (n=33). Signal intensities of PI DNA stain (red) and FISH signal (green) were measured along the length of the chromosomes and binned according to their position, from the centromere (0-5%) to the telomere (95-100%), before plotting. Error bars represent +/- standard error from the mean (SEM). Average DNA stain intensity was compared between the region of 50-70% (*S. pombe* DNA corresponding to the highest FISH signal) and either 25-45% or 75-95% (endogenous mouse DNA) by the KS test (*p<0.01, n.s. = not significant).

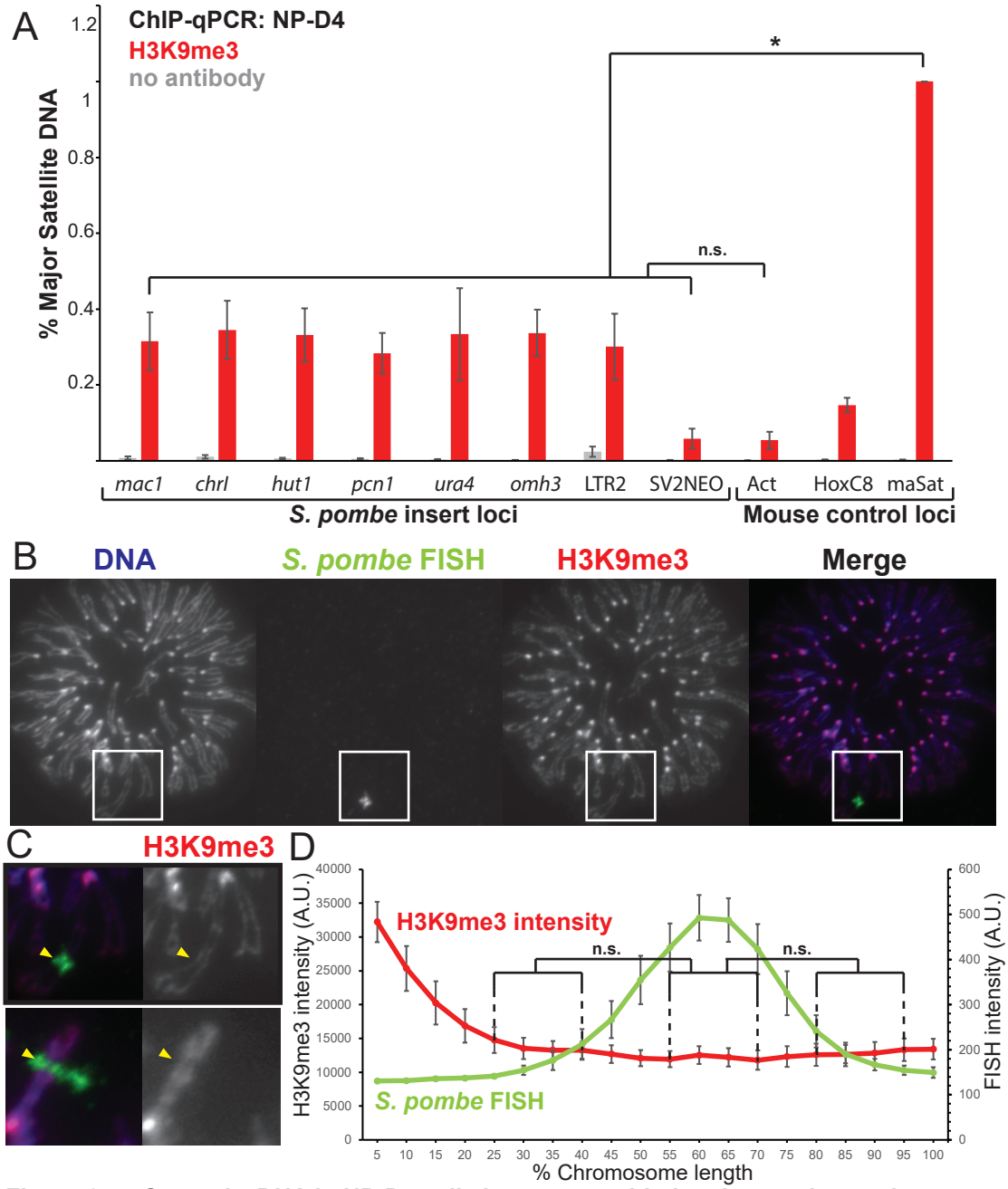


**Figure 6.7 - *S. pombe* DNA in NP-D4 cells is not assembled on heterochromatin**

**A** - ChIP-qPCR of NP-D4 interphase cells for H3K9me3 at 8 loci within the *S. pombe* insert and 3 mouse control loci. Act is the highly transcribed negative control gene actin, HoxC8 is a region of facultative heterochromatin and maSat is the centromeric satellite positive control region. Error bars represent +/- SEM of three independent repeats. Enrichments were normalised to maSat levels and compared by the *t*-test (*p<0.05, n.s. = not significant).
**B & C** - Immunolocalisation for H3K9me3 (red channel) on NP-D4 metaphase spreads. The *S. pombe* insert region is visualised by FISH (green channel). DNA is stained with DAPI (blue channel). The site of *S. pombe* DNA is indicated in **C** by yellow arrows.
**D** - Average chromosome profile for H3K9me3 on the NP-D4 insert bearing chromosome across several images (n=19). Immunolocalisation (red) and FISH (green) signal intensity were measured along the length of the chromosomes and binned according to their position. Error bars represent +/- SEM. Average H3K9me3 intensity was compared between the region of 55-70% (*S. pombe* DNA corresponding to the highest FISH signal) and either 25-40% or 80-95% (endogenous mouse DNA) by the KS test (*p<0.01, n.s. = not significant).

## 6.5 Discussion

The goal of this chapter was to determine if H3K9me3 heterochromatin affects mitotic chromosome structure as observed at *S. pombe* DNA inserted in mammalian chromosomes. Initial attempts to remove H3K9me3 heterochromatin from the *S. pombe* insert in F1.1 cells by inducing a pluripotent state were unsuccessful. Neither standard iPS reprogramming through expression of the reprogramming factors Oct4, Sox2, cMyc and Klf4, nor fusion of F1.1 with ESCs induced a pluripotent state in the target cell (**Figure 6.1**). Some cell types have been reported to be more resistant to iPS reprogramming than others, and most studies have thus instead focused on cell lines shown to be amenable to the standard methods (Sakurada 2010, Malik and Rao 2013). We could find no instance of C127 cells being reprogrammed in the literature and our results suggest that these cells may not be among those able to be reprogrammed with current techniques either by expression of the iPS reprogramming factors or by cell fusion. It is of note that unlike most somatic cell lines, F1.1 / C127 cells grew well in ESC media. This suggests that C127 cells may already share some properties with stem cells that prevents them from fully responding to somatic reprogramming protocols. The continued and rapid growth of C127 cells during the sensitive reprogramming period may also hinder the formation and isolation of stem cell-like colonies. For most somatic cell lines, the growth in ESC media of only those cells which have begun to reprogram allows them to form colonies without competition and overcrowding from the original somatic cells. The lack of space for colony expansion and the necessity to passage the entire population might prevent the formation of fully reprogrammed colonies. The growth of the F1.1 / C127 cells in ESC media could therefore pose a significant obstacle to the reprogramming of these cell lines.

Attempts to synthetically remove H3K9me3 from the *S. pombe* insert in F1.1 cells were only partially successful. Expression of the Suv39CD-EYFP-JMJD2D construct, designed to bind to and remove any H3K9me3 marks *in vivo*, only reduced H3K9me3 over the *S. pombe* DNA by 50% (**Figure 6.4 A**). While this represents a considerable decrease, it is not sufficient to confidently relate to any change or lack thereof in mitotic chromosome structure. Indeed, our main criterion for satisfactory removal of H3K9me3 should be that the *S. pombe* chromatin becomes indistinguishable from mouse chromatin with respect to H3K9me3 levels.

Suv39CD-EYFP-JMJD2D expressing cells retained a distinct H3K9me3 pattern on the *S. pombe* insert relative to surrounding normal chromatin (**Figure 6.4 B**). Heterochromatin was therefore insufficiently reduced to allow for conclusions to be made.

It remains to be determined why the PREditOR system removes H3K9me3 so efficiently from centromeres but incompletely from the *S. pombe* insert. The answer to this question may lie in the choice of chromodomain used to target H3K9 demethylation. The chromodomain of Suv39H1 is known to be a general binder of H3K9me3 (Wang *et al.* 2012). However, the binding activity and targeting of the Suv39 proteins and the other mammalian H3K9 KMTs, SetDB1, G9a and GLP, is complex and not fully understood. Suv39H1 and Suv39H2 are the primary H3K9 KMTs responsible for H3K9me at centromere repeats (Peters *et al.* 2001, Rice *et al.* 2003), likely explaining the efficient targeting of PREditOR demethylation to major satellites. While the chromodomain of Suv39H1 and Suv39H2 does bind both H3K9me3 and HP1 (Melcher *et al.* 2000), this likely acts in concert with other factors to efficiently target these proteins to centromere repeats (Krouwels *et al.* 2005). For instance, the binding of Suv39H1 to major satellite RNA through its CD was recently shown to be important for Suv39H1 recruitment and H3K9me3 deposition at centromere repeats (Johnson *et al.* 2017, Shirai *et al.* 2017). Thus, while the general binding of the Suv39H1 CD to H3K9me3 can explain the partial removal of H3K9me3 at non-centromeric regions such as the *S. pombe* DNA, the lack of additional factors for the stable recruitment of the Suv39 CD to the *S. pombe* DNA may result in less efficient targeting of PREditOR to this region and thus less complete demethylation compared to major satellites. This outcome highlights the fact that the identity of the H3K9 KMT responsible for the high levels of H3K9me3 over the *S. pombe* insert in F1.1 cells remains unknown. The other KMTs, notably SetDB1, are believed to be more important in the deposition of H3K9me3 at regions outside the centromeres (Kim and Kim 2012, Yeap *et al.* 2009, Yuan *et al.* 2009). The fusion of a demethylase to a catalytically inactive SetDB1 might therefore have been a more appropriate construct for the purposes of this experiment. Alternatively, the chromodomain of HP1, which should be recruited strongly to all regions of H3K9me3 heterochromatin, might achieve better targeting (Bannister *et al.* 2001, Lachner *et al.* 2001).

However, insertion of a large region of *S. pombe* DNA into a mammalian chromosome without H3K9me3 was achieved by transfecting naked DNA rather than fusion with *S. pombe* protoplasts. This resulted in the NP-D4 cell line, in which the inserted *S. pombe* DNA was indistinguishable from surrounding chromatin in terms of both its mitotic chromosome structure and associated H3K9me3 levels (**Figures 6.5-6.7**). In a first instance, this suggests that the DNA sequence is not the basis for the formation of either heterochromatin or a distinct mitotic structure over *S. pombe* DNA in mammalian cells. While the composition of the *S. pombe* DNA inserted into F1.1, NP-A4, HeP-3 and NP-D4 must certainly vary considerably, all are large regions of rearranged sequence derived from the same, relatively small genome. Substantial parts of the *S. pombe* inserts are therefore bound to be shared between the cell lines, and indeed PCR analyses show that a number of loci are shared between all of them. Presumably, they also share the same broad sequence features, including GC-content, that are properties of the *S. pombe* genome.

This therefore suggests that both the high H3K9me3 levels and the distinct mitotic structure of the *S. pombe* insert, whether related to each other or not, result from some feature associated with *S. pombe* chromatin introduced upon protoplast fusion but not by DNA transfection. The model put forward in **Chapter 5**, whereby H3K9me3 heterochromatin spreads from regions of pre-existing heterochromatin within the *S. pombe* genome, would explain the requirement of introducing *S. pombe* chromatin for the formation of heterochromatin over the *S. pombe* insert in the mammalian host cell. An effect of heterochromatin on the organisation of mitotic chromatin into loops might then link the observed distinct structure to heterochromatin spreading. In this way, the correlation of H3K9me3 heterochromatin and an altered chromatin structure over *S. pombe* DNA in the fusion derived cells only would be explained.

The low level of H3K9me3 detected on *S. pombe* DNA in NP-D4 is certainly more compelling than the partial reduction observed in the Suv39CD-EYFP-JMJD2D expressing F1.1 cell line. While not as low as on active genes, the H3K9me3 levels over *S. pombe* DNA in NP-D4 are similar to those associated with the intergenic host genome regions surrounding the F1.1 insert, albeit in a different mouse cell line (**Figures 6.7 A & 5.4**). Immunolocalisation on mitotic chromosomes in NP-D4 confirmed that H3K9me3 levels are indistinguishable between the *S. pombe* insert

and the surrounding mouse chromatin (**Figure 6.7 B-D**). These observations thus fulfil our previously mentioned criterion for an acceptable decrease in H3K9me3 levels.

In a DNA transfection-derived cell line, neither H3K9me3 heterochromatin nor a distinct structure were associated with the inserted *S. pombe* DNA. Combined with the correlation of H3K9me3 heterochromatin and a distinct structure over the inserts of the F1.1, NP-A4 and HeP-3 cells, this strongly suggests a link between these two features of chromatin. Other factors, including the cell type and the source of the inserted DNA, are common between at least some of the cell lines, leaving the high level of H3K9me3 over the *S. pombe* insert as the only common feature present in all cells where the distinct structure is observed and absent where it is not.

# Chapter 7: Discussion

## 7.1 Introduction

The exact structure of the mitotic chromosome and the mechanisms by which it is assembled and regulated are the subjects of much current research. According to the radial loop model, chromatin is organised into loops emanating out from a central scaffold of organiser proteins, which include the important condensin complexes (Maeshima and Eltsov 2008, Hirano 2012, Gibcus *et al.* 2018). These complexes are proposed to form these loops by extruding chromatin through their ring structures (Golobordko *et al.* 2016, Gibcus *et al.* 2018). While this model explains much of the data, there remain many gaps in it that need to be filled. The place in this model of several proteins with a known role in mitotic chromosome structure, such as TopoIIα and Kif4, has not yet been determined (Chang *et al.* 2003, Samejima *et al.* 2012, Kchonsak and Haering 2015). The role of factors other than organising proteins in these processes, such as histone PTMs, is not yet fully understood either (Zhiteneva *et al.* 2017). Underlying all of these uncertainties is the major question of how models of mitotic chromosome structure might account for variability in structure, as is observed, for instance, between the centromere and chromosome arms (Blackburn and Szostak 1984).

To explore this question, this study investigated regions of distinct chromatin structure assembled on foreign *S. pombe* DNA inserted into mammalian chromosomes. Beginning with the previously described F1.1 cell line (Allshire *et al.* 1987, McManus *et al.* 1994), the structure of its insert was analysed and found to be consistent with an altered mitotic organisation into smaller loops compared to surrounding chromatin. Two other cell lines were then generated by fusion in which a similar distinct structure assembled over *S. pombe* chromatin. In all of these cell lines H3K9me3 heterochromatin was found to be enriched over the *S. pombe* DNA. Generation of a cell line by DNA transfection yielded an *S. pombe* insert that lacked both heterochromatin and a distinct structure. This suggests a model whereby H3K9me3 heterochromatin can alter the structure of the mitotic chromosome by either directly or indirectly affecting the activity of the condensin complexes. This chapter will discuss these results and the possibilities and implications of this model.

## 7.2 *S. pombe* DNA stably inserted into mammalian cells is highly rearranged

Early analyses of the DNA content of the F1.1 cell line suggested that it contained several Mb of rearranged *S. pombe* DNA (McManus *et al.* 1994). The results presented in **Chapter 3** confirmed this hypothesis, although the DNA was perhaps found to be more rearranged than initially anticipated. While several large and potentially intact regions of several hundred kb of *S. pombe* DNA were identified by sequencing, other regions as small as 30 or 40 kb were also found to be present in F1.1. A substantial proportion of the *S. pombe* regions inserted into the F1.1 genome was also found to have done so in two copies.

For the purposes of analysing the *S. pombe* insert in detail, the insertion of a large, contiguous and single copy region of DNA would be most desirable. However, the insertion of intact Mb size regions of DNA into a mammalian genome is not easily achieved, if at all possible. Mammalian cells have long been known to rearrange introduced DNA (Wigler *et al.* 1978, Perucho *et al.* 1980). While the sequence analysis of the other mammalian – *S. pombe* cell lines generated here was not as in depth as for F1.1, the state of the *S. pombe* DNA in these cells appears to be similar. NP-A4, HeP-2, HeP-3 and NP-D4 were all found to contain sequences from all three *S. pombe* chromosomes, as illustrated by the loci analysed in the ChIP-qPCR results presented in **Chapters 5 & 6**. Whether by cell fusion or transfection, the insertion of *S. pombe* sequences into mammalian cells, either mouse or human, resulted in substantial rearrangement of the sequence prior to insertion. Thus, the insertion of an intact region of DNA at a scale large enough to visualise in mitosis is not feasible with current methods.

## 7.3 *S. pombe* chromatin in mammalian – *S. pombe* fusion cells is organised into smaller loops in mitosis

### 7.3.1 The structure of *S. pombe* chromatin in mammalian cells fits with radial loop models of mitotic chromosome structure

Several advances have led the radial loop model of mitotic chromosome structure to gain precedence over other models in recent years (Naumova *et al.* 2009, Gibcus *et*

*al.* 2018). The appearance and analysis of the *S. pombe* chromatin in the F1.1, NP-A4 and HeP-3 cell lines presented in **Chapter 4** fits well with the modern conception of this model, lending it further credence. The decreased amount of chromatin over the *S. pombe* insert seen by microscopy and the restricted range of DNA interactions seen by Hi-C are both readily explained by a different local organisation of chromatin into smaller loops. In a radial loop model, a smaller loop organisation would manifest itself as a narrower stretch of chromosome, which is exactly what is observed and how the structure was initially described in F1.1 (McManus *et al.* 1994).

Other models of mitotic chromosome structure are difficult to reconcile with these observations. In a hierarchical folding model, the mitotic chromosome results from the folding of chromatin into increasingly larger fibres (Belmont *et al*., 1987, Woodcock and Gosh 2010). The size of the chromatin fibres thus increases stepwise, with each step representing a significant increase in diameter. In order to account for a local decrease in size of the chromosome, this model would therefore require a partial unfolding of the chromatin fibre in the vicinity of the *S. pombe* DNA. Aside from the difficulties of achieving this mechanistically, even a partial unfolding would result in a decrease in size of the fibre of at least 2-3 fold, far more than the modest decrease in diameter observed. Hierarchical folding models are thus unsuited to situations involving a minor alteration of structure.

The only other way that the hierarchical folding model or the network model (Poirier and Marko 2002) could account for a region of narrower diameter on the chromosome would be to somehow locally compact the chromatin more. This explanation would be more feasible as such compaction could occur at a lower level of organisation, for instance by increasing the interaction between nucleosomes, without altering the larger scale organisation. However, such compaction would also be at odds with the observed decrease in the amount of chromatin over the *S. pombe* insert (**Figures 4.1, 4.4 & 4.5**). An increase in the compaction of chromatin would be expected to increase the amount of chromatin contained within a certain volume or length of chromosome. Conversely, the *S. pombe* chromatin appears less compact than surrounding chromatin, which is again compatible with an organisation into smaller loops in the radial loop model.

The best explanation is therefore that chromatin is organised into loops around an organising scaffold which remains constant while the loops can vary. How this constant scaffold / variable loop organisation is achieved is another question.

### 7.3.2 The loop extrusion model predicts an unchanging scaffold of condensin

While the radial loop model has long been one of the main models for the nature of the structure of the mitotic chromosome, it is the loop extrusion model that has most recently been put forward to explain the mechanism by which that structure is formed (Goloborodko *et al.* 2016, Gibcus *et al.* 2018). According to this model, condensins can form a radial loop chromosome structure simply by binding to chromatin and pushing chromatin loops through their rings. Altering certain parameters of the condensins can alter how the chromatin is organised relative to the condensin. However, the structure of the scaffold should be primarily determined by the interaction between condensin complexes as they assemble along the linear chromosome. The nature of that interaction is simple and unchanging: one condensin complex cannot extrude chromatin beyond the barrier posed by another condensin complex. The model could therefore largely be said to predict that regardless of the final structure of the chromatin, the organisation of the condensin complexes relative to each other, i.e. the structure of the scaffold, should not change.

This prediction proved to hold true in the F1.1 cell line, which showed no change in overall condensin levels between the regions of different chromatin structure assembled on mouse or *S. pombe* DNA. This observation therefore lends further support to an active role for condensin in mitotic chromosome formation consistent with a combined loop extrusion / radial loop model. If condensin complexes acted simply as a “glue” holding chromatin fibres together (Kireeva *et al.* 2004) they would be expected to scale with the amount of chromatin. The consistency of the condensin levels over the chromosome arms therefore points to it being a driving force in the organisation of both chromatin and itself. However, this is in contrast to centromeres, where condensin levels are found to scale with chromatin concentration, both of them increasing (**Figures 4.1, 4.9 & 4.10**). Loop extrusion alone is therefore not sufficient to explain the organisation of the scaffold at

centromeres, although it does fit our observations of the chromosome arms, including the region of *S. pombe* chromatin.

### 7.3.3 Condensin activity may be locally altered to change chromatin loop size

If the organisation of the condensin scaffold remains constant, then alteration of the parameters of condensin-chromatin interaction must necessarily alter the structure of the chromatin around the scaffold. An organisation into shorter chromatin loops could be achieved by increasing the loading of condensin onto the chromatin fibre, or by decreasing the processivity of its loop extrusion activity (**Figure 4.14**). It is worth noting that a change in the parameters of only one condensin complex but not the other would be equivalent to a change in the condensin ratio. As just discussed, according to the loop extrusion model, the organisation of the chromosome scaffold is determined by the interaction between condensin complexes. While this leads us to expect total condensin to remain constant, the ratio of condensin I to condensin II within that unchanging total can be altered by a change in their relative activity. Given that no change in that ratio was detected in F1.1 cells (**Figures 4.9 & 4.10**), it must be concluded that whatever the change in condensin activity that is leading to the shorter loop organisation of the *S. pombe* chromatin is, it must be affecting condensin I and condensin II equally. This is not difficult to imagine. Condensin I and II share the SMC2 and SMC4 proteins that form their rings, and their cap proteins are homologous (Hirano 2012). The activity of each complex must be different, as attested by their different effects on chromosome morphology which has been attributed to a different loop size that each mediates (Shintomi and Hirano 2011, Gibcus *et al.* 2018). However, a factor that affects both complexes in a similar manner, for instance by impeding the ATPase activity of the shared SMC subunits, is highly plausible.

While theoretically condensin activity could be globally altered in this way to produce different chromosome morphologies in different cells, the structural change discussed here only affects a portion of one chromosome. Whatever is altering condensin activity must therefore be localised to that region of chromatin. An inherent property of the *S. pombe* chromatin, or a tightly associated factor, that modifies condensin activity is thus the most likely explanation for its altered mitotic structure. Heterochromatin fits this criterion as a distinct property of the chromatin

associated with the *S. pombe* DNA. The mechanism by which heterochromatin could influence condensin activity will be discussed later.

### 7.3.4 Cohesin may be directly involved in mitotic chromosome structure

Alternatively, cohesin may be playing a role in the structure of the *S. pombe* insert. Being in the same family as condensin, the cohesin complex is an important organiser of chromatin (Nasmyth and Haering 2009). Its primary role in mitosis is in the cohesion of sister chromatids, however a secondary role in mitotic chromosome structure has been postulated and proposed as an explanation for the distinct appearance of centromeres (Hirota *et al.* 2004). This structural role has usually been attributed to the effect of cohesin on the ratio of condensin I to condensin II, as the presence of cohesin has been proposed to hinder the loading of condensin II (Shintomi and Hirano 2011). The finding that cohesin persists over the *S. pombe* chromatin in a significant proportion of F1.1 cells raises the possibility that it may be contributing to the distinct structure of that chromosomal region. This possibility is made more compelling by the observation that the number of F1.1 insert bearing chromosomes which show cohesin persistence and the number which have a distinct structure are very close, about 70% in each case, although these numbers were obtained in different experiments.

However, the unchanging ratio of condensin I to condensin II over the *S. pombe* chromatin precludes the possibility of cohesin influencing chromosome structure through an alteration of that ratio. If a role for cohesin in the structure of the *S. pombe* insert is to be admitted two possibilities remain:

1. Cohesin could be acting through condensin in a way that does not alter the condensin I to II ratio. In this situation the distinct structure of the *S. pombe* chromatin is still due to a change in condensin activity as previously discussed, but that change in activity is caused by the presence of cohesin. As a highly similar complex to condensin, it is possible that the presence of cohesin could block the extrusion activity of condensin in a similar way to that in which another condensin complex would. Thus, though cohesin is not necessarily involved in loop formation itself, it could lead to the formation of smaller loops by blocking the progress of both

condensin I and II, preventing their processing of large loops. In this way, it would mimic the situation in which condensin loading was increased.

2. Cohesin could have a role in mitotic chromosome structure separate from condensin. In this situation cohesin would be actively involved in loop formation, for instance by the formation of additional loops-within-loops on top of those already formed by condensin I and II. Such a role has never been attributed to cohesin in mitosis, where it is primarily thought to act by holding together separate strands of chromatin rather than adjacent ones in a loop. However, the loop extrusion model is a primary mechanism proposed for the organisation of chromatin by cohesin in interphase, leaving at least the possibility that this activity might be maintained in mitosis (Sanborn *et al.* 2015).

The activity of cohesin is another mechanism by which the change in chromatin structure could be traced back to heterochromatin. As will be discussed in more detail later, the persistence of cohesin at the *S. pombe* chromatin could be a result of heterochromatin protecting it from degradation, as is the case at centromeres (Bernard and Allshire 2002).

### 7.3.5 A role for other chromatin organising proteins?

In trying to explain the appearance of regions of distinct chromatin architecture, this study has primarily focused on the condensin proteins as the best described mitotic chromatin organising proteins. In doing so, a potential role for cohesin has also presented itself. Further study might determine if there is any role for other mitotic chromosome architecture proteins such as TopoIIα and Kif4. The role of these proteins is less well described and is thought to be secondary with respect to the more central condensin complexes, as well as being more important in the later stages of mitosis (Samejima *et al.* 2012, Kchonsak and Haering 2015). Nonetheless, examination of their activity at the inserted *S. pombe* chromatin would be an interesting avenue for future research and a further potential source of structural variation in chromosomes.

## 7.4 H3K9me3 spreads from *S. pombe* heterochromatin to cover the inserted *S. pombe* chromatin in mammalian cells

### 7.4.1 Heterochromatin in the mammalian host cell spreads from pre-existing *S. pombe* heterochromatin

The insertion of chromatin from *S. pombe* into mouse or human cells by cell fusion was shown to lead to the establishment of a large domain of H3K9me3 heterochromatin over the *S. pombe* DNA. Conversely, *S. pombe* DNA from the same source inserted by transfection, without chromatin proteins, was assembled into unremarkable chromatin resembling bulk chromatin (**Chapters 5 & 6**). This suggests that the heterochromatin originates from some property of the *S. pombe* chromatin proteins that is transmitted to the mammalian chromatin. No other potential sources of information can account for all the observations. The DNA in all cell lines is from the same source and thus highly similar. Cell type variability cannot account for the differences as heterochromatin assembles on the *S. pombe* insert after fusion with several cell types and species, mouse C127 and NIH-3T3 and human HeLa, but fails to do so in transfected NIH-3T3 cells. While the exact insertion site is not known for the new cell lines, there are no features of the F1.1 insertion site which suggest that it would confer heterochromatic identity to the *S. pombe* insert. On the contrary, the *S. pombe* DNA was inserted in close proximity to two active genes on either side of the insertion site and heterochromatin was not found to have spread beyond the *S. pombe* chromatin.

The most plausible explanation is therefore that heterochromatin is spreading from within the *S. pombe* chromatin to establish a large domain over the entirety of the insert region after its insertion in the mammalian genome (**Figure 5.8**). As an apparently random assembly of regions from the *S. pombe* genome, any insert of several Mb in size is highly likely to contain regions of constitutive heterochromatin from the *S. pombe* genome, such as centromeres, telomeres or silenced mating type loci (Allshire and Ekwall 2015). Indeed, in the F1.1 cell line, for which the sequence is known, several such regions are known to be present in the 12 Mb insert (**Figure 3.1**). The rearrangement of the *S. pombe* sequence also means that these heterochromatic regions will be evenly spread out across the insert.

For heterochromatin to spread from these regions to cover the whole insert, two things are required:

1. The heterochromatin spreading machinery of the mammalian host cell must be able to recognise and propagate the heterochromatin signal of the *S. pombe* chromatin. The most well-described mechanism for the spreading of heterochromatin involves the recognition of the H3K9me3 mark by proteins such as Suv39H1/H2 and HP1, which can then propagate H3K9me3 to adjacent nucleosomes in cis by a positive feedback loop (Allshire and Madhani 2017). All that is required for this mechanism to act on the *S. pombe* chromatin is for these proteins to recognise their binding and substrate targets, which in both cases are histones. Histones and their modifications are highly conserved across eukaryotes, with for instance 92% identity between histone H3 in mouse and *S. pombe* (Postberg *et al.* 2010). The recognition of these proteins by the machinery of other species therefore seems likely.

2. The spreading of heterochromatin must not be hindered across the *S. pombe* sequence. This too is a highly likely outcome. Unlike for highly conserved histone proteins, the genomes of mammalian and yeast are highly divergent. Sequence elements which might act as a barrier to heterochromatin spreading, such as insulator binding sites or regions of transcription, are unlikely to be recognised between species. Indeed, transcription from the *S. pombe* insert appears to be completely shut down, removing one of the main barriers to heterochromatin spreading (**Figure 5.5**). On the other hand, features of the endogenous genome would halt the spreading of heterochromatin quite well, which explains why H3K9me3 does not appear to have extended into the surrounding regions of chromatin.

### 7.4.2 When are *S. pombe* histones replaced by mammalian histones after fusion?

An interesting question is to ask at what point the inserted *S. pombe* chromatin is converted to the chromatin of the host cell. Although the *S. pombe* DNA is introduced to the mammalian cell assembled on purely *S. pombe* chromatin, the *S. pombe* histones will at some point be replaced by mammalian histones. This raises

numerous questions about how and when the H3K9me3 is spread. Are the histones replaced before or after the DNA is integrated into the host genome? Does the H3K9me3 propagate over the insert on *S. pombe* histones prior to the switch or on mammalian histones after it? Does the switch occur independent of or linked to replication of the *S. pombe* DNA? Mechanisms exist which have been shown to transmit histone PTMs, including H3K9me3, through replication (Rivera *et al.* 2014, Audergon *et al.* 2015). None of these situations therefore pose any major problems to the heterochromatin spreading hypothesis. However, determining the exact order of events after introduction of the chromatin would be interesting.

### 7.4.3 H3K9me3 is established prior to 5meC over the *S. pombe* DNA

Also interesting would be to establish the link in these cells between H3K9me3 and 5meC. In mammals these marks exist in a mutually reinforcing and often interdependent state (Rose and Klose 2014). *S. pombe* on the other hand does not possess DNA methylation (Capuano *et al.* 2014). Our analysis of the NP-A4 and HeP-3 cells suggests that H3K9me3 is established over the *S. pombe* chromatin reasonably quickly after fusion, while 5meC was previously shown to increase gradually over many generations on the F1.1 insert (McManus *et al.* 1994). In this case then, H3K9me3 seems to be the first mark to be established, which then leads to a gradual, and perhaps more robust, silencing by 5meC. Whether the H3K9me3 mark can propagate across cell generations in the absence of 5meC in mammals, i.e. whether H3K9me3 has true “epigenetic” properties, has been a longstanding question (Zhu and Reinberg 2011). If H3K9me3 is indeed established prior to 5meC in the mammalian – *S. pombe* fusion cells, it would be strong evidence for the epigenetic propagation of H3K9me3 independent of DNA methylation.

### 7.4.4 A model for the establishment of H3K9me3 heterochromatin over *S. pombe* DNA in mammalian cells

Our model is therefore that regions of heterochromatin from *S. pombe* act as a platform from which H3K9me3 spreads in the mammalian host by the action of host proteins recognising *S. pombe* H3K9me3, but unhindered by any barrier domains within the insert sequence. In this way heterochromatin is spread over the *S. pombe* chromatin in fusion cells, but not in transfected cells, in which the information carried

by the *S. pombe* chromatin is lost (**Figure 7.1**). An interesting test of this model would be to fuse mammalian cells with *S. pombe* cells lacking H3K9me3. This can be achieved by deletion of the sole *S. pombe* H3K9 KMT *clr4* (Nakayama *et al.* 2001). These *S. pombe* cells would possess all the same chromatin features as wild type *S. pombe* apart from H3K9me, and would therefore not be expected to be able to spread heterochromatin from any region within their genome. Recreation of the fusion experiments with this strain would thus rule out any possibility of the heterochromatin establishment in fusion cells resulting from some property of the *S. pombe* chromatin other than H3K9me3.

## 7.5 Heterochromatin alters the organisation of chromatin in mitosis

### 7.5.1 High levels of H3K9me3 correlate with a distinct structure over *S. pombe* DNA

The results presented in **Chapters 5 & 6** laid out the evidence for a causative link between H3K9me3 heterochromatin and the distinct mitotic structure that assembles over *S. pombe* DNA in mammalian – *S. pombe* fusion cells. These two properties of the chromatin were found to be correlated in all three fusion cells, including both mouse and human cells, and the distinct structure was absent in cells in which the DNA was introduced by transfection and did not form heterochromatin. While the association is compelling, useful future work would expand the range of these observations to further examples and other cell lines to strengthen the correlation. The experiment just described in **Section 7.4.4** to test heterochromatin spreading would also provide additional support for this hypothesis. Fusion of a mammalian cell with a *clr4Δ* strain of *S. pombe* would be expected to yield a non-heterochromatic insert of *S. pombe* chromatin. Based on this hypothesis, we would therefore not expect the insert to adopt a distinct structure in mitosis. This experiment would therefore confirm that the distinct structure observed in fusion cells is due to the presence of heterochromatin, and not to some other result of the fusion procedure not achieved by transfection.

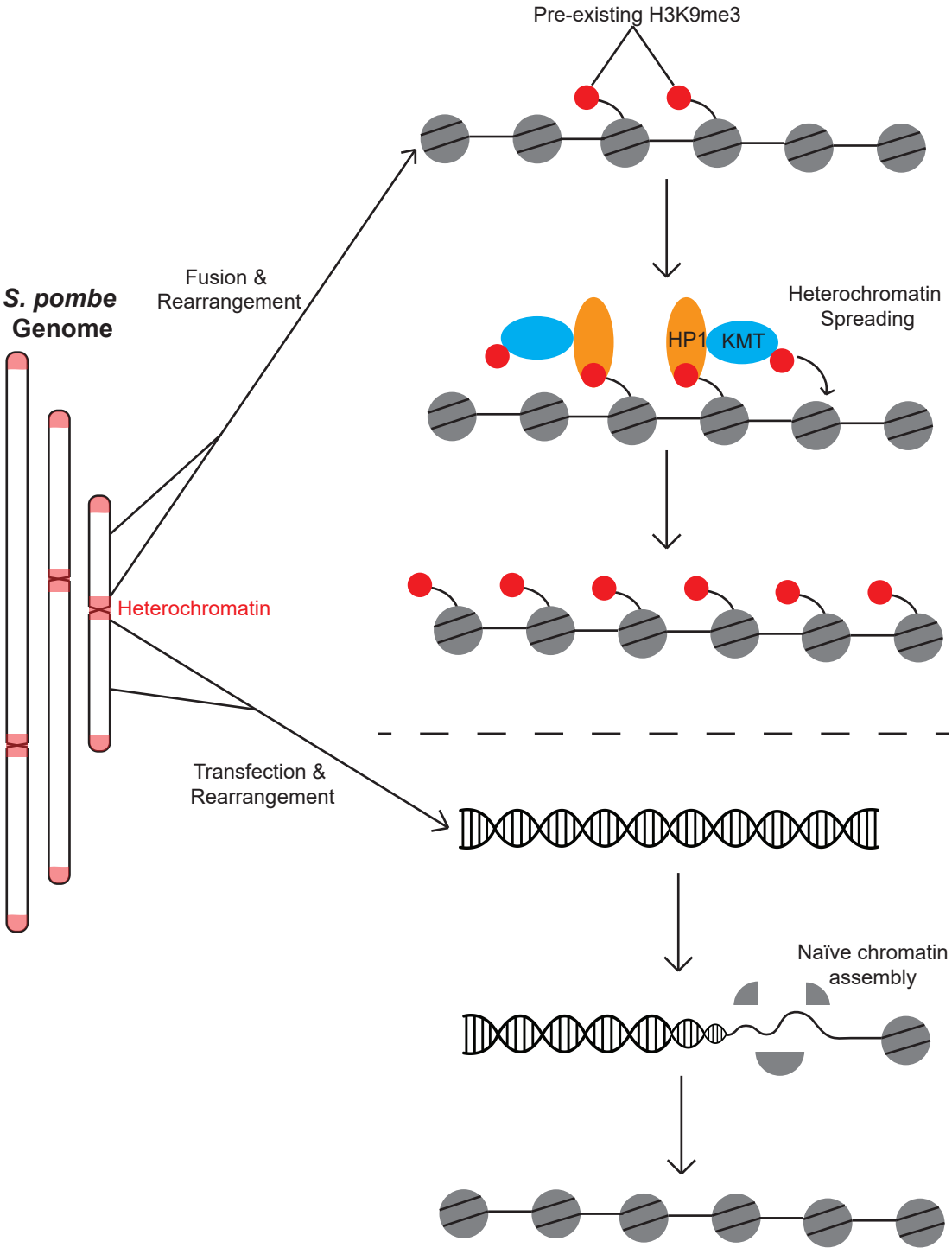


**Figure 7.1 - Heterochromatin spreads from pre-existing *S. pombe* heterochromatin in mammalian fusion cells but not DNA transfectants**

In fusion cells (top) *S. pombe* DNA is inserted as chromatin and certain regions maintain their heterochromatic identity from the *S. pombe* cell. This heterochromatin is recognised and propagated by the mammalian proteins to establish a heterochromatic domain over the *S. pombe* DNA. In DNA transfected cells (bottom) *S. pombe* DNA is introduced as naked DNA and loses all heterochromatic identity. It is then assembled naïvely on mammalian nucleosomes and thus does not form heterochromatin.

### 7.5.2 H3K9me3 and chromatin compaction

A role for histone PTMs in regulating the structure of chromatin is well known. This role is best described at the lowest level of chromatin organisation, the 10nm fibre, where PTMs modulate the interactions between nucleosomes, DNA and other proteins (Bannister and Kouzarides 2011). In addition, histone PTMs have been proposed to be a major contributor to mitotic chromosome structure, and have even been proposed to lead to "compaction" of chromatin by processes independent of the "shaping" action of organising proteins such as the condensin complexes (Zhiteneva *et al.* 2017). This mitotic role for PTMs may well be related to their functions in interphase and may involve similar processes of mediating nucleosome interactions. H3K9me3 in particular has been shown to increase during mitosis, and a role for it in chromosome compaction has been put forward (McManus *et al.* 2006, Park *et al.* 2011). This mitotic role may involve similar processes to those that lead to compaction of H3K9me3 heterochromatin in interphase, although any role for HP1 must be discounted as it dissociates from chromatin in mitosis (Danzer and Wallrath 2004, Fischle *et al.* 2005, Hirota *et al.* 2005).

However, most work so far has focused on finding PTMs which vary globally across chromatin in a cell cycle dependent manner, contributing to the overall architecture of chromatin in mitosis. A difference in a particular PTM over different regions of chromatin leading to variation in structure has not yet been described. Furthermore, the results presented here support a role for H3K9me3 in regulating chromosome "shaping" by condensins or other proteins rather than "compaction". This would therefore represent a new role for H3K9me3 heterochromatin in mitosis.

### 7.5.3 Heterochromatin may directly alter the activity of condensin complexes

The hypothesis put forward here is that H3K9me3 heterochromatin leads to the organisation of chromatin into smaller loops by altering the activity of the condensin complexes. The simplest way this might be achieved is by a direct effect of H3K9me3 on condensin processivity. This effect may be related to its ability to compact chromatin fibres. One possibility is that the ability of condensin to extrude loops of chromatin is hindered by chromatin compaction. Alternatively, heterochromatin could be interpreted as a signal for the increased loading of

condensin and recognised either directly by condensin or indirectly by an associated loading factor. Biochemical analysis of the affinity and processing activity of condensin on heterochromatin vs. euchromatin would be an interesting first step in dissecting the mechanism by which H3K9me3 influences condensin activity.

### 7.5.4 Heterochromatin may alter chromatin structure through protection of cohesin

A further interesting relationship to investigate would be that between heterochromatin and cohesin at the *S. pombe* insert. The potential effect of cohesin on chromosome structure, either directly or through its effect on condensin, has already been discussed. The effect of heterochromatin on mitotic chromosome structure may thus be mediated through its effect on cohesin. H3K9me3 is an essential component of the pathway by which cohesin is protected from degradation at pericentromeres until anaphase onset (Bernard and Allshire 2002). It seems highly likely then that the persistence of cohesin over the *S. pombe* insert region in F1.1 is a result of the high H3K9me3 levels being interpreted as a pericentromeric signal. Given this well-established link between heterochromatin and cohesin, the most plausible scenario is that the alteration of mitotic chromosome structure by heterochromatin is an indirect effect achieved through the action of cohesin.

The model put forward in **Figure 7.2** summarises this process. Condensin loading and the early stages of compaction proceed identically on all types of chromatin and in the presence of cohesin, which in early prophase is present on chromosome arms. In later prophase cohesin is removed from most of the chromosome arms but is protected and persists at H3K9me3 heterochromatin. Cohesin inhibits the loop extrusion of condensin by impeding its progress along the chromatin fibre, leading to the formation of smaller chromatin loops over heterochromatic regions than over non-heterochromatic regions. If this occurs over a large enough region, such as over the *S. pombe* DNA in fusion cells, the mitotic structure of the chromosome becomes visibly altered locally over the heterochromatin.

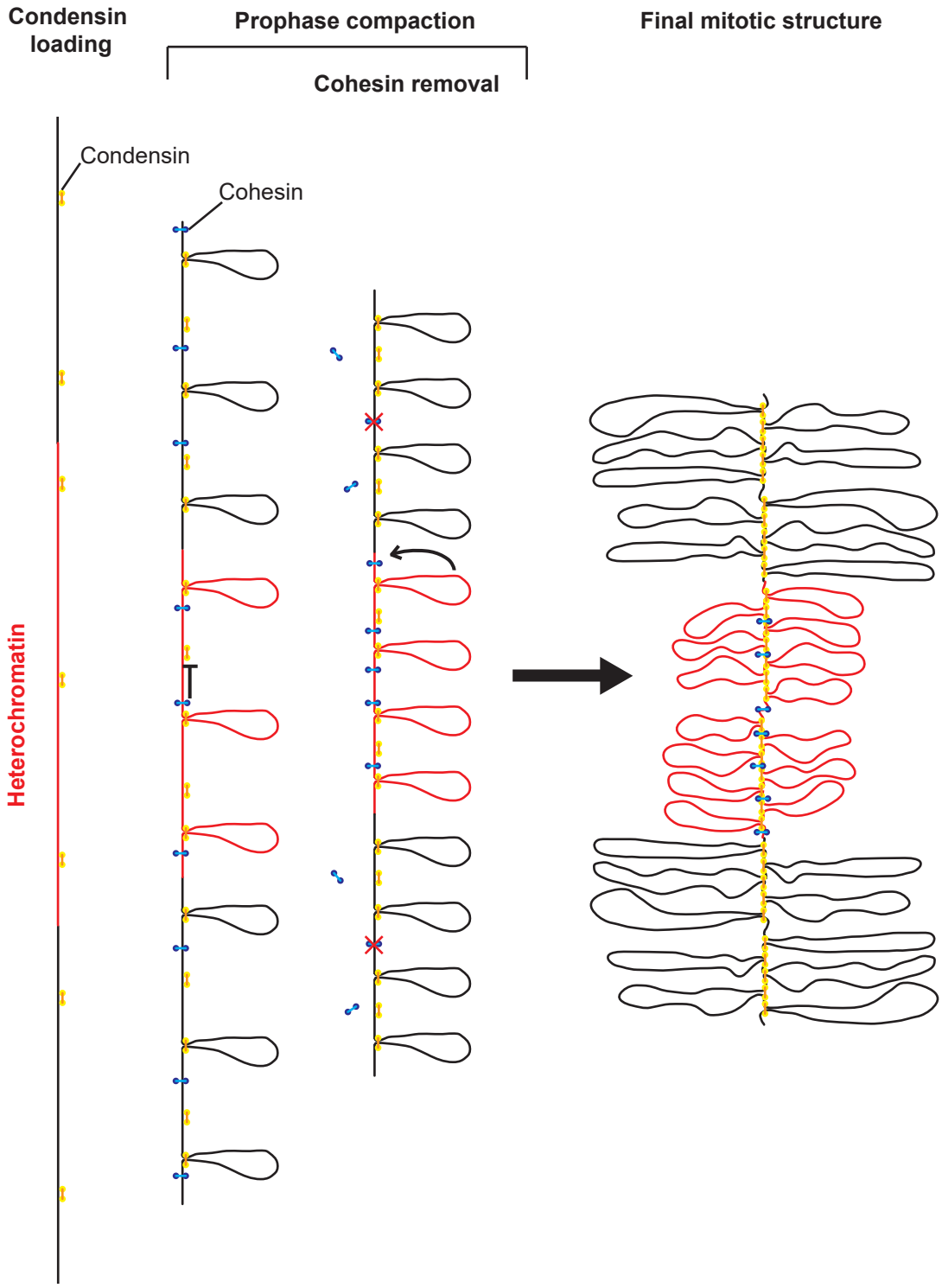


**Figure 7.2 - Model: Heterochromatin is organised into smaller loops in mitosis by retaining cohesin, which inhibits condensin activity**

Early stages of chromatin compaction proceed similarly for heterochromatin and non-heterochromatin. Cohesin inhibits the activity of condensin by blocking loop extrusion. In late prophase cohesin is removed from the chromosome arms except where it is protected by heterochromatin. Cohesin continues to inhibit loop extrusion at heterochromatin only, leading to an organisation into smaller loops over the regions where it persists.

## 7.6 Can heterochromatin explain the structure of the centromere?

A major question posed by this model is whether the process described here is altering the structure of the chromatin in naturally occurring heterochromatin, notably at the centromere. If heterochromatin is altering chromatin organisation, particularly if this is achieved through cohesin activity, we would expect it to act similarly at centromeres. Centromeres have long been known to appear as constrictions on chromosomes, a structure which might be consistent with an organisation in smaller loops as presented here. However, other aspects of centromere appearance differ from that of the *S. pombe* inserts. In the mouse cell lines F1.1 and NP-A4 in particular, centromeres appear as regions of bright DNA staining in stark contrast to the darkened regions of *S. pombe* DNA (**Figures 4.1 & 4.6**). On the other hand, the *S. pombe* insert in the human HeP-3 cell line is extremely similar in appearance on the chromosome to the centromere (**Figure 4.8**).

Comparisons in mouse cells must be qualified by the caveat that centromeres of mouse chromosomes are telocentric, thereby combining the properties of centromeres and telomeres. Further comparisons in these and other species could perhaps determine with more certainty if certain structural properties are shared between the regions of H3K9me3 heterochromatin over *S. pombe* DNA and centromeric regions. It certainly seems likely that the effect described here at *S. pombe* chromatin will also be relevant at centromeres, given their similarly high H3K9me3 levels. However, this effect may be combined or in competition with others at centromeres, which have a highly distinctive epigenetic profile beyond simply H3K9me3 and where a number of complex processes related to kinetochore function are at work (Verdaasdonk and Bloom 2011). The analysis of the *S. pombe* inserts in F1.1, NP-A4 and HeP-3 may simply have allowed the specific effect of heterochromatin on mitotic chromosome structure to be isolated and examined over a large region of heterochromatin, assembled on unique sequence and without confounding factors related to centromere function. The influence of H3K9me3 heterochromatin on mitotic chromosome structure is thus one of potentially many factors that contribute to the distinct structure of centromeres in mitosis.

## 7.7 Conclusions and perspectives

The organisation of chromatin in mitosis is a complex process involving a large number of related players, many of which have yet to be elucidated. Through examination of foreign DNA inserted into mammalian chromosomes, this study identified a potential role for H3K9me3 heterochromatin as a regulator of mitotic chromosome structure through its effect on the important organising condensin complexes. H3K9me3 is an essential component of centromeric regions in many eukaryotes and is known to fulfil several roles. These results may therefore have wide ranging implications for the structure of centromeres in these species. It may also imply that histone post-translational modifications play a more important role in mitotic chromosome organisation, and in particular local variations in chromatin structure, than has previously been supposed. Future work could identify others among the many PTMs which have hitherto unknown roles in regulating the structure of mitotic chromosomes.

## Acknowledgements

I would like to thank my supervisor Robin for unearthing a great project for me from the depths of the MRC cryostore that let me do something a bit different and turned out to be a lot more than just “molecular archaeology”. I would like to thank him also for his constant help and approachability over the past four years as well as his guidance and patience during my writing up period, with its interruptions for post-doc interviews and unexpected trips to ITV studios, Salford. The whole Allshire lab have been amazing right from the start and very helpful in any way they could to the one weird non-yeast scientist in their midst. Thank you to Sharon and Alison for providing a calming and delightful day to day presence, and Manu and Ryan for doing everything they could to make it chaotic. Likewise, the constant Yin and Yang between Pauline’s sunny disposition and what can only be described as Sandra’s stormy presence was always fun. Thank you to Pin for the bioinformatics help for someone who prefers to use computers for video games, to Tanya for stories of her never-ending adventures at conferences, to Sito for having someone to share both cheese and dessert with, to Puneet for having someone else to blame when our shared bin was overflowing, and to Desi, Roberta and Tadgh for relieving the pressure of being the only person to not write gene names lower case and italicised. Then there are the people across town I knew I could turn to when I needed to know something about mammalian cells. Heidi and Richard for their help with the stem cells and general cell culture, Shelagh for showing me how to do FISH and helping me troubleshoot when I then went on to use labelled RNA instead of DNA by accident for a few months, and Nick Gilbert for showing a genuine interest in my project and exchanging ideas. Job Dekker, Hakan Ozadam and Johann Gibcus have been indispensable for the Hi-C and Ana Losada provided me with some very good antibodies. Finally, I would like to thank my family and partner Fiona for their love and support. Luckily for them I think as PhD students go I am on the more optimistic side, although perhaps that is due in no small part to them.

## References


Adachi Y, Käs E & Laemmli UK (1989) Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. *The EMBO journal* **8:** 3997–4006

Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J & Bazett-Jones DP (2010) Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS ONE* **5:**

Aizawa H, Sekine Y, Takemura R, Zhang Z, Nangaku M & Hirokawa N (1992) Kinesin family in murine central nervous system. *Journal of Cell Biology* **119:** 1287–1296

Alipour E & Marko JF (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Research* **40:** 11202–11212

Allshire RC & Ekwall K (2015) Epigenetics regulation of chromatin states in *Schizosaccharomyces pombe*. *Cold Spring Harb Perspectives in Biology* **7:** 1–25

Allshire RC, Cranston G, Gosden JR, Maule JC, Hastie ND & Fantes PA (1987) A fission yeast chromosome can replicate autonomously in mouse cells. *Cell* **50:** 391–403

Allshire RC & Madhani HD (2017) Ten Principles of Heterochromatin Formation and Function. *Nature Publishing Group***:** 1–48

Antonin W & Neumann H (2016) Chromosome condensation and decondensation during mitosis. *Current Opinion in Cell Biology* **40:** 15–22

Babu KA & Verma RS (1985) Structural and functional aspects of nucleolar organizer regions (NORs) of human chromosomes. *International Review of Cytology* **94:** 151–176

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC & Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410:** 120–124

Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell research* **21:** 381–395

Baxter J, Sen N, Lopez Martinez V, Monturus de Carandini ME, Schvartzman JB, Diffley JFX & Aragon L (2011) Positive Supercoiling of Mitotic DNA Drives Decatenation by Topoisomerase II in Eukaryotes. *Science* **311:** 518–522

Becker JS, Nicetto D & Zaret KS (2016) H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends in Genetics* **32:** 29–41

Bell O, Tiwari VK, Thomä NH & Schübeler D (2011) Determinants and dynamics of genome accessibility. *Nature Reviews Genetics* **12:** 554–564

Belmont AS (2006) Mitotic chromosome structure and condensation. *Current opinion in cell biology* **18:** 632–8

Belmont AS, Sedat JW & Agard DA (1987) A three-dimensional approach to mitotic chromosome structure: evidence for a complex hierarchical organization.

Bemer M & Grossniklaus U (2012) Dynamic regulation of Polycomb group activity during plant development. *Current Opinion in Plant Biology* **15:** 523–529

Bernard P, Maure JF, Partridge JF, Genier S, Javerzat JP & Allshire RC (2001) Requirement of heterochromatin for cohesion at centromeres. *Science* **294:** 2539–2542

Bernard P & Allshire RC (2002) Centromeres become unstuck without heterochromatin. *Trends in Cell Biology* **12:** 419–424

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL & Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–26

Bickmore WA & Van Steensel B (2013) Genome architecture: Domain organization of interphase chromosomes. *Cell* **152:** 1270–1284

Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* **16:** 6–21

Bird AP (1987) CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics* **3:** 342–347

Blackburn EH & Szostak JW (1984) The molecular structure of centromeres and teleomeres. *Annual Review of Biochemistry* **53:** 163–194

Blackledge NP, Rose NR & Klose RJ (2015) Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nature Reviews Molecular Cell Biology* **16:** 643–649

Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K & Meissner A (2011) Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144:** 439–452

Boyle S, Rodesch MJ, Halvensleben HA, Jeddeloh JA & Bickmore WA (2011) Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Research* **19:** 901–909

Brockdorff N (2013) Noncoding RNA and Polycomb recruitment. *Rna* **19:** 429–442

Busslinger G a., Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N & Peters J-M (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544:** 503–507

Campanero MR, Armstrong MI & Flemington EK (2000) CpG methylation as a mechanism for the regulation of E2F activity. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 6481–6486

Campos EI & Reinberg D (2009) Histones: annotating chromatin. *Annual review of genetics* **43:** 559–99

Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ & Al-sady B (2011) Chromodomain-Mediated Oligomerization of HP1

Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Molecular Cell* **41:** 67–81

Capuano F, Mülleder M, Kok R, Blom HJ & Ralser M (2014) Cytosine DNA Methylation Is Found in *Drosophila melanogaster* but Absent in *Saccharomyces cerevisiae, Schizosaccharomyces pombe* and Other Yeast Species. *Analytical Chemistry* **86:** 3697–3702

Carlberg C & Molnár F (2014) *Mechanisms of Gene Regulation,* Springer

Caspersson T, Lomakka G & Zech L (1971) The 24 fluorescence patterns of the human metaphase chromosomes - distinguishing character and variability. *Hereditas* **102:** 89–102

Chang C-J, Goulding S, Earnshaw WC & Carmena M (2003) RNAi analysis reveals an unexpected role for topoisomerase II in chromosome arm congression to a metaphase plate. *Journal of Cell Science* **116:** 4715–26

Chen T & Dent SYR (2014) Chromatin modifiers and remodellers: Regulators of cellular differentiation. *Nature Reviews Genetics* **15:** 93–106

Clapier CR & Cairns BR (2009) The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry* **78:** 273–304

Coelho P a, Queiroz-Machado J & Sunkel CE (2003) Condensin-dependent localisation of topoisomerase II to an axial chromosomal structure is required for sister chromatid resolution during mitosis. *Journal of Cell Science* **116:** 4763–76

Cowan CA, Atienza J, Melton DA & Eggan K (2005) Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells. *Science* **309:** 1369–1373

Cremer T & Cremer M (2010) Chromosome territories. *Cold Spring Harbor perspectives in biology*

Cuellar TL, Herzner A-M, Zhang X, Goyal Y, Watanabe C, Friedman BA, Janakiraman V, Durinck S, Stinson J, Arnott D, Cheung TK, Chaudhuri S, Modrusan Z, Doerr JM, Classon M & Haley B (2017) Silencing of retrotransposons by SETDB1

inhibits the interferon response in acute myeloid leukemia. *The Journal of Cell Biology* **216:** 3535–3549

Danzer JR & Wallrath LL (2004) Mechanisms of HP1-mediated gene silencing in Drosophila. *Development* **131:** 3571–3580

Daujat S, Zeissler U, Waldmann T, Happel N & Schneider R (2005) HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *Journal of Biological Chemistry* **280:** 38090–38095

Day JJ & Sweatt JD (2010) DNA methylation and memory formation. *Nature Neuroscience* **13:** 1319–1323

De Koning L, Corpet A, Haber JE & Almouzni G (2007) Histone chaperones: An escort network regulating histone traffic. *Nature Structural and Molecular Biology* **14:** 997–1007

Demmerle J, Koch AJ & Holaska JM (2013) Emerin and histone deacetylase 3 (HDAC3) cooperatively regulate expression and nuclear positions of MyoD, Myf5, and Pax7 genes during myogenesis. *Chromosome Research* **21:** 765–779

Dimitrova E, Turberfield a. H & Klose RJ (2015) Histone demethylases in chromatin biology and beyond. *EMBO reports* **16:** 1620–1639

DiNardo S, Voelkel K & Sternglanz R (1984) DNA topoisomerase II mutant of Saccharomyces cerevisiae: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **81:** 2616–20

Disteche CM (1999) Escapees on the X chromosome. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 14180–2

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS & Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–80

Dobigny G, Ozouf-Costaz C, Bonillo C & Volobouev V (2004) Viability of X-autosome translocations in mammals: An epigenomic hypothesis from a rodent case-study. *Chromosoma* **113:** 34–41

Dong KB, Maksakova IA, Mohn F, Leung D, Appanah R, Lee S, Yang HW, Lam LL, Mager DL, Schübeler D, Tachibana M, Shinkai Y & Lorincz MC (2008) DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO Journal* **27:** 2691–2701

Dubochet J, Adrian M, Schultz P & Oudet P (1986) Cryo-electron microscopy of vitrified SV40 minichromosomes: the liquid drop model. *The EMBO journal* **5:** 519–28

Durkin SG & Glover TW (2007) Chromosome Fragile Sites. *Annual Review of Genetics* **41:** 169–192

Duthie SM, Nesterova TB, Formstone EJ, Keohane AM, Turner BM, Zakian SM & Brockdorff N (1999) Xist RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Human Molecular Genetics* **8:** 195–204

Earnshaw WC, Halligan B, Cooke CA, Heck MMS & Liu LF (1985) Topoisomerase II is a structural componenet of mitotic chromosome scaffolds. *Jcb* **100:** 1706–1715

Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RDG, Buetow KH, Gingeras TR, Misteli T & Meshorer E (2008) Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2:** 437–447

Eissenberg JC & Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental Biology* **339:** 240–249

Elgin SCR & Reuter G (2013) Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology* **5:** 1–26

Eltsov M, Maclellan KM, Maeshima K, Frangakis AS & Dubochet J (2008) Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 19732–19737

Entrevan M, Schuettengruber B & Cavalli G (2016) Regulation of Genome Architecture and Function by Polycomb Proteins. *Trends in Cell Biology* **26:** 511–525

Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultchi AI, Wutz A & Bickmore W a (2010) Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Molecular cell* **38:** 452–64

Espada J, Ballestar E, Fraga MF, Villar-Garea A, Juarranz A, Stockert JC, Robertson KD, Fuks F & Esteller M (2004) Human DNA methyltransferase 1 is required for maintenance of the histone H3 modification pattern. *Journal of Biological Chemistry* **279:** 37175–37184

Farcas AM, Blackledge NP, Sudbery I, Long HK, McGouran JF, Rose NR, Lee S, Sims D, Cerase A, Sheahan TW, Koseki H, Brockdorff N, Ponting CP, Kessler BM & Klose RJ (2012) KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **1:** e00205

Fatemi M & Wade P a (2006) MBD family proteins: reading the epigenetic code. *Journal of Cell Science* **119:** 3033–3037

Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nature reviews. Genetics* **12:** 565–75

Finch JT & Klug a (1976) Solenoidal model for superstructure in chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **73:** 1897–1901

Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR & Bickmore WA (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genetics* **4:**

Fischle W, Boo ST, Dormann HL, Ueberheide BM, Garcia BA, Shabanowitz J, Hunt DF, Funabiki H & Allis CD (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **438:** 1116–1122

Flor-Parra I, Zhurinsky J, Bernal M, Gallardo P & Daga RR (2009) A Lallzyme MMX-based rapid method for fission yeast protoplast preparation. *Yeast (Chichester, England)* **26:** 545–551

Folger KIMR, Wong EA, Wahl G, Capecchil MR & Capecchi MR (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Molecular and cellular biology* **2:** 1372–87

Foshay KM, Looney TJ, Chari S, Mao FF, Lee JH, Zhang L, Fernandes CJ, Baker SW, Clift KL, Gaetz J, Di C-G, Xiang AP & Lahn BT (2012) Embryonic stem cells induce pluripotency in somatic cell fusion through biphasic reprogramming. *Molecular cell* **46:** 159–70

Francis NJ, Kingston RE & Woodcock CL (2004) Chromatin compaction by a polycomb group protein complex. *Science* **306:** 1574–7

Fuks F, Hurd PJ, Deplus R & Kouzarides T (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Research* **31:** 2305–2312

Fussner E, Ching RW & Bazett-Jones DP (2011) Living without 30nm chromatin fibers. *Trends in biochemical sciences* **36:** 1–6

Gan ES, Xu Y & Ito T (2015) Dynamics of H3K27me3 methylation and demethylation in plant development. *Plant Signaling and Behavior* **10:**

Gasser SM, Laroche T, Falquet J, Boy de la Tour E & Laemmli UK (1986) Metaphase chromosome structure. Involvement of topoisomerase II. *Journal of Molecular Biology* **188:** 613–629

Gasser SM & Laemmli UK (1986) Cohabitation of Scaffold Binding Regions with Upstream / Enhancer Elements of Three Developmentally Regulated Genes of *D. melanogaster*. **46:** 521–530

Gibcus JH, Samejima K, Goloborodko A, Samejima I, Naumova N, Nuebler J, Kanemaki MT, Xie L, Paulson JR, Earnshaw WC, Mirny LA & Dekker J (2018) A pathway for mitotic chromosome formation. *Science* **6135:** eaao6135

Goloborodko A, Marko JF & Mirny LA (2016) Chromosome Compaction by Active Loop Extrusion. *Biophysical Journal* **110:** 2162–2168

Goto H, Yasui Y, Nigg EA & Inagaki M (2002) Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes to Cells* **7:** 11–17

Grewal S & Jia S (2007) Heterochromatin revisited. *Nature Reviews Genetics* **8:** 35–46

Groth A, Rocha W, Verreault A & Almouzni G (2007) Chromatin Challenges during DNA Replication and Repair. *Cell* **128:** 721–733

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, De Klein A, Wessels L, De Laat W & Van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951

Gurley LR, Walters RA & Tobey RA (1975) Sequential phosphorylation of histone subfractions in the chinese hamster cell cycle. *Journal of Biological Chemistry* **250:** 3936–3944

Han J & Sidhu KS (2011) Reprogramming Somatic Cells by Fusion with Embryonic Stem Cells : Present Status and Prospects in Regenerative Medicine. *Methodological Advances in the Culture, Manipulation and Utilization of Embryonic Stem Cells for Basic and Practical Applications*

Happel N & Doenecke D (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431:** 1–12

Hauf S, Waizenegger IC, Peters J, Natl P, Hauf S, Waizenegger IC & Peters J-M (2001) Cleavage by Separase Required for Anaphase and Cytokinesis in Human Cells. **293:** 1320–1323

Haushalter KA & Kadonaga JT (2003) Chromatin assembly by DNA-translocating motors. *Nature Reviews Molecular Cell Biology* **4:** 613–620

Heard E, Chaumeil J, Masui O & Okamoto I (2004) Mammalian X-chromosome inactivation: An epigenetics paradigm. *Cold Spring Harbor Symposia on Quantitative Biology* **69:** 89–102

Henikoff S & Smith MM (2015) Histone variants and epigenetics. *Cold Spring Harbor Perspectives in Biology* **7:** 1–26

Hennig L & Derkacheva M (2009) Diversity of Polycomb group complexes in plants: same rules, different players? *Trends in Genetics* **25:** 414–423

Hickman MA, Froyd CA & Rusche LN (2011) Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. *Eukaryotic Cell* **10:** 1183–1192

Hirano T (2012) Condensins: Universal organizers of chromosomes with diverse functions. *Genes and Development* **26:** 1659–1678

Hirota T, Gerlich D, Koch B, Ellenberg J & Peters J-M (2004) Distinct functions of condensin I and II in mitotic chromosome assembly. *Journal of cell science* **117:** 6435–45

Hirota T, Lipp JJ, Toh BH & Peters JM (2005) Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438:** 1176–1180

Holm C, Goto T, Wang JC & Botstein D (1985) DNA topoisomerase II is required at the time of mitosis in yeast. *Cell* **41:** 553–563

Holoch D, Moazed D & Avenue L (2015) RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics* **16:** 71–84

Hong L, Schroth GP, Matthews HR, Yau P & Bradbury EM (1993) Studies of the DNA binding properties of histone H4 amino terminus. *Journal of Biological Chemistry* **268:** 305–314

Horn PJ, Carruthers LM, Logie C, Hill DA, Solomon MJ, Wade PA, Imbalzano AN, Hansen JC & Peterson CL (2002) Phosphorylation of linker histones regulates ATP-depenent chromatin remodeling enzymes. *Nature Structural Biology* **9:** 263–267

HOROWITZ RA, AGARD DA, SEDAT JW & Woodcock CL (1994) The 3-Dimensional Architecture of Chromatin in-Situ - Electron Tomography Reveals Fibers Composed of a Continuously Variable Zigzag Nucleosomal Ribbon. *Journal of Cell Biology* **125:** 1–10

Hsieh L-J, Cheng Y-M, Wang Y-C, Lin C-C & Li Y-C (2014) Organization and evolution of a novel cervid satellite DNA with yeast CDEI-like repeats. *Zoological Studies* **53:** 25

Huang C, Wang X, Liu X, Cao S & Shan G (2015) RNAi pathway participates in chromosome segregation in mammalian cells. *Cell Discovery* **1:** 15029

Hudson DF, Vagnarelli P, Gassmann R & Earnshaw WC (2003) Condensin is required for nonhistone protein assembly and structural integrity of vertebrate mitotic chromosomes. *Developmental cell* **5:** 323–36

Iguchi-Ariga SMM & Schaffner W (1989) CpG methylation of the cAMP- responsive enhancer / promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes & Development***:** 612–619

Isono K, Endo TA, Ku M, Yamada D, Suzuki R, Sharif J, Ishikura T, Toyoda T, Bernstein BE & Koseki H (2013) SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Developmental Cell* **26:** 565–577

Izzo A & Schneider R (2016) The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1859:** 486–495

Jeltsch A. (2006) Molecular Enzymology of Mammalian DNA Methyltransferases. In: Doerfler W., Böhm P. (eds) *DNA Methylation: Basic Mechanisms. Current Topics in Microbiology and Immunology* **301** Springer

Jenuwein T & Allis CD (2001) Translating the Histone Code. *Science* **293:** 1074–1080

Jiang Y, Matevossian A, Guo Y & Akbarian S (2011) Setdb1-mediated histone H3K9 hypermethylation in neurons worsens the neurological phenotype of Mecp2-deficient mice. *Neuropharmacology* **60:** 1088–1097

Johnson WL, Yewdell WT, Bell JC, Mcnulty S, Duda Z, Neill RJO, Sullivan B a & Straight AF (2017) Molecular Cell RNA dependent stabilization of SUV39H1 at constitutive heterochromatin. *eLife***:** 1–32

Kellum R & Alberts BM (1995) Heterochromatin protein 1 is required for correct chromosome segregation in Drosophila embryos. *Journal of cell science* **108 ( Pt 4:** 1419–1431

Kim DH & Sung S (2012) Environmentally coordinated epigenetic silencing of FLC by protein and long noncoding RNA components. *Current Opinion in Plant Biology* **15:** 51–56

Kireeva N, Lakonishok M, Kireev I, Hirano T & Belmont AS (2004) Visualization of early chromosome condensation: A hierarchical folding, axial glue model of chromosome structure. *Journal of Cell Biology* **166:** 775–785

Klose RJ, Kallin EM & Zhang Y (2006) JmjC-domain-containing proteins and histone demethylation. *Nature reviews. Genetics* **7:** 715–27

Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE & Meissner A (2011) Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell* **8:** 96–105

Korf BR & Diacumakos EG (1978) Microsurgically-extracted metaphase chromosomes of the Indian Muntjac examined with phase contrast and scannign electron microscopy. *Experimental cell research* **1:**

Kornberg RD & Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98:** 285–294

Kouzarides T (2007) Chromatin Modifications and Their Function. *Cell* **128:** 693–705

Krouwels IM, Wiesmeijer K, Abraham TE, Molenaar C, Verwoerd NP, Tanke HJ & Dirks RW (2005) A glue for heterochromatin maintenance: Stable SUV39H1 binding to heterochromatin is reinforced by the SET domain. *Journal of Cell Biology* **170:** 537–549

Krude T (1995) Chromatin: Nucleosome assembly during DNA replication. *Current Biology* **5:** 1232–1234

Kschonsak M & Haering CH (2015) Shaping mitotic chromosomes: From classical concepts to molecular mechanisms. *BioEssays* **37:** 755–766

Kubista M, Aakerman B & Norden B (1987) Characterization of interaction between DNA and 4',6-diamidino-2-phenylindole by optical spectroscopy. *Biochemistry* **26:** 4545–4553

Lachner M, O'Carroll D, Rea S, Mechtler K & Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410:** 116–20

Lai SK, Wong CH, Lee YP & Li HY (2011) Caspase-3-mediated degradation of condensin Cap-H regulates mitotic cell death. *Cell Death and Differentiation* **18:** 996–1004

Laird A, Thomson JP, Harrison DJ & Meehan RR (2013) 5-Hydroxymethylcytosine Profiling As an Indicator of Cellular State. *Epigenomics* **5:** 655–69

Langmore JP & Paulson JR (1983) Low angle x-ray diffraction studies of chromatin structure in vivo and in isolated nuclei and metaphase chromosomes. *The Journal of cell biology* **96:** 1120–1131

Lee YM, Lee S, Lee E, Shin H, Hahn H, Choi W, Kim W (2001) Human kinesin superfamily member 4 is dominantly localized to the nuclear matrix and is associated with chromosomes during mitosis. *Biochemical Journal* **360:** 549-556

Leeb M & Wutz A (2012) Establishment of epigenetic patterns in development. *Chromosoma* **121:** 251–262

Lehnertz B, Ueda Y, Derijck AAHA, Braunschweig U, Perez-burgos L, Kubicek S, Chen T, Li E, Jenuwein T & Peters AHFM (2003) Suv39h -Mediated Histone H3

Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. **13:** 1192–1200

Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P, Li Y, Szulwach KE, Jin P, Lorincz MC & Ren B (2014) Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proceedings of the National Academy of Sciences* **111:** 6690–6695

Lewis CD & Laemmli UK (1982) Higher order metaphase chromosome structure: evidence for metalloprotein interactions. *Cell* **29:** 171–181

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender M a, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny L a, Lander ES & Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289–93

Lomberk G, Wallrath LL & Urrutia R (2006) The Heterochromatin Protein 1 family. *Genome Biology* **7:** 1–8

Luger K, Mäder AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:** 251–260

Maeshima K & Eltsov M (2008) Packaging the genome: the structure of mitotic chromosomes. *Journal of biochemistry* **143:** 145–53

Maeshima K, Imai R, Tamura S & Nozaki T (2014) Chromatin as dynamic 10-nm fibers. *Chromosoma* **123:** 225–37

Maeshima K & Laemmli UK (2003) A two-step scaffolding model for mitotic chromosome assembly. *Developmental cell* **4:** 467–80

Maison C, Quivy JP, Probst A V. & Almouzni G (2010) Heterochromatin at mouse pericentromeres a model for *de novo* heterochromatin formation and duplication during replication. *Cold Spring Harbor Symposia on Quantitative Biology* **75:** 155–165

Malik N & Rao M (2013) A Review of the Methods for Human iPSC Derivation. *Methods Molecular Biology* **997:** 23–33

Malone CD & Hannon GJ (2009) Small RNAs as Guardians of the Genome. *Cell* **136:** 656–668

Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD & Reinberg D (2008) Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular cell* **32:** 503–18

Margueron R & Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469:** 343–9

Marko JF (2008) Micromechanical studies of mitotic chromosomes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **16:** 469–97

Martin RM, Leonhardt H & Cardoso MC (2005) DNA labeling in living cells. *Cytometry Part A* **67:** 45–52

Matarese F, Carrillo-de Santa Pau E & Stunnenberg HG (2011) 5-Hydroxymethylcytosine: a new kid on the epigenetic block? *Molecular systems biology* **7:** 562

Mattout A, Biran A & Meshorer E (2011) Global epigenetic changes during somatic cell reprogramming to iPS cells. *Journal of molecular cell biology* **3:** 341–50

Matzke MA & Mosher RA (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nature Reviews Genetics* **15:** 394–408

McManus J, Perry P, Sumner a T, Wright DM, Thomson EJ, Allshire RC, Hastie ND & Bickmore WA (1994) Unusual chromosome structure of fission yeast DNA in mouse cells. *Journal of cell science* **107:** 469–86

McManus KJ, Biron VL, Heit R, Underhill DA & Hendzel MJ (2006) Dynamic changes in histone H3 lysine 9 methylations: Identification of a mitosis-specific function for dynamic methylation in chromosome congression and segregation. *Journal of Biological Chemistry* **281:** 8888–8897

Melcher M, Schmid M, Aagaard L, Selenko P, Laible G & Jenuwein T (2000) Structure-Function Analysis of SUV39H1 Reveals a Dominant Role in Heterochromatin Organization, Chromosome Segregation, and Mitotic Progression. *Molecular and Cellular Biology* **20:** 3728–3741

Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT & Misteli T (2006) Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell* **10:** 105–116

Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES & Meissner A (2008) Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454:** 49–55

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–60

Mirkovitch J, Gasser SM & Laemmli UK (1988) Scaffold attachment of DNA loops in metaphase chromosomes. *Journal of Molecular Biology* **200:** 101–109

Molina O, Carmena M, Maudlin IE & Earnshaw WC (2016) PREditOR: a synthetic biology approach to removing heterochromatin from cells. *Chromosome Research***:** 1–15

Monier K, Mouradian S & Sullivan KF (2006) DNA methylation promotes Aurora-B-driven phosphorylation of histone H3 in chromosomal subdomains. *Journal of Cell Science* **120:** 101–114

Mora-Bermúdez F, Gerlich D & Ellenberg J (2007) Maximal chromosome compaction occurs by axial shortening in anaphase and depends on Aurora kinase. *Nature cell biology* **9:** 822–31

Moreno S, Klar A & Nurse P (1991) Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. *Methods in Enzymology* **194:** 795–823

Morgan HD, Santos F, Green K, Dean W & Reik W (2005) Epigenetic reprogramming in mammals. *Human Molecular Genetics* **14:** 47–58

Murray, M G and Thompsom W (1980) Nucleic Acids Research. *Nucleic Acids Research* **8:** 4321–4326

Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y & Nakano T (2012) PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486:** 415–419

Nakayama J, Rice JC, Strahl BD, Allis CD & Grewal SIS (2001) Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* **292:** 110–113

Nasmyth K (2002) Segregating sister genomes: The molecular biology of chromosome separation. *Science* **297:** 559–565

Nasmyth K & Haering CH (2009) Cohesin: Its Roles and Mechanisms. *Annual Review of Genetics* **43:** 525–558

Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny L a & Dekker J (2013) Organization of the mitotic chromosome. *Science* **342:** 948–53

Nguyen CT, Weisenberger DJ, Velicescu M, Gonzales FA, Lin JCY, Liang G & Jones PA (2002) Histone H3-Lysine 9 Methylation Is Associated with Aberrant Gene Silencing in Cancer Cells and Is Rapidly Reversed by Advances in Brief Histone H3-Lysine 9 Methylation Is Associated with Aberrant Gene Silencing in Cancer Cells and Is Rapidly Reversed by 5. *Cancer Research* **62:** 6456–6461

Nishino Y, Eltsov M, Joti Y, Ito K, Takata H, Takahashi Y, Hihara S, Frangakis AS, Imamoto N, Ishikawa T & Maeshima K (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO journal* **31:** 1644–53

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, Van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J & Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485:** 381–385

Ohta S, Bukowski-Wills JC, Sanchez-Pulido L, Alves F de L, Wood L, Chen ZA, Platani M, Fischer L, Hudson DF, Ponting CP, Fukagawa T, Earnshaw WC & Rappsilber J (2010) The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. *Cell* **142:** 810–821

Okano M, Bell DW, Haber DA & Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99:** 247–257

Okita K, Ichisaka T & Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. *Nature* **448:** 313–317

Ono T, Losada A, Hirano M, Myers MP, Neuwald AF & Hirano T (2003) Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115:** 109–21

Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, Cobb BS, Yokomori K, Dillon N, Aragon L, Fisher AG & Merkenschlager M (2008) Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **132:** 422–433

Park J-A, Kim A-J, Kang Y, Jung Y-J, Kim HK & Kim K-C (2011) Deacetylation and methylation at histone H3 lysine 9 (H3K9) coordinate chromosome condensation during cell cycle progression. *Molecules and Cells* **31:** 343–349

Passarge E (1979) Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *American journal of human genetics* **31:** 106–15

Paulson JF & Laemmli UK (1977) Chromosomes of Histone-Depleted Metaphase. *Cell* **12:** 817–828

Perucho M, Hanahan D & Wigler M (1980) Genetic and physical linkage of exogenous sequences in transformed cells. *Cell* **22:** 309–317

Peters AHFM, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, Opravil S, Doyle M, Sibilia

M & Jenuwein T (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107:** 323–337

Phillips JE & Corces VG (2009) CTCF: Master Weaver of the Genome. *Cell* **137:** 1194–1211

Pollard TD & Earnshaw WC (2002) *Cell Biology*, Elsevier

Postberg J, Forcob S, Chang WJ & Lipps HJ (2010) The evolutionary history of histone H3 suggests a deep eukaryotic root of chromatin modifying mechanisms. *BMC Evolutionary Biology* **10:** 1–13

Prigent C & Dimitrov S (2003) Phosphorylation of serine 10 in histone H3, what for? *Journal of Cell Science* **116:** 3677–3685

Qian C & Zhou MM (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cellular and Molecular Life Sciences* **63:** 2755–2763

Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M & Cosma MP (2015) Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160:** 1145–1158

Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, Shinkai Y & Allis CD (2003) Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. *Molecular Cell* **12:** 1591–1598

Rivera C, Gurard-Levin ZA, Almouzni G & Loyola A (2014) Histone lysine methylation and chromatin replication. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1839:** 1433–1439

Roque A, Ponte I, Arrondo JLR & Suau P (2008) Phosphorylation of the carboxy-terminal domain of histone H1: Effects on secondary structure and DNA condensation. *Nucleic Acids Research* **36:** 4719–4726

Rose NR & Klose RJ (2014) Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1839:** 1362–1372

Routh A, Sandin S & Rhodes D (2008) Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proceedings of the National Academy of Sciences* **105:** 8872–8877

Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA & Krumm A (2008) CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 8309–14

Ruvinsky A & Graves JM (2004) *Mammalian Genomics,* CABI

Saha A, Wittmeyer J & Cairns BR (2006) Chromatin remodelling: The industrial revolution of DNA around histones. *Nature Reviews Molecular Cell Biology* **7:** 437–447

Saka Y, Sutani T, Yamashita Y, Saitoh S, Takeuchi M, Nakaseko Y & Yanagida M (1994) Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis. *The EMBO journal* **13:** 4938–52

Sakurada K (2010) Environmental epigenetic modifications and reprogramming-recalcitrant genes. *Stem Cell Research* **4:** 157–164

Samejima K, Samejima I, Vagnarelli P, Ogawa H, Vargiu G, Kelly D a, de Lima Alves F, Kerr A, Green LC, Hudson DF, Ohta S, Cooke C a, Farr CJ, Rappsilber J & Earnshaw WC (2012) Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase IIα. *The Journal of cell biology* **199:** 755–70

Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES & Aiden EL (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences* **112:** 201518552

Sarraf SA & Stancheva I (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Molecular Cell* **15:** 595–605

Schalch T, Duda S, Sargent DF & Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436:** 138–141

Schmiesing J, Gregson H, Zhou S & Yokomori K (2000) A human condensin complex containing hCAP-C-hCAP-E and CNAP1, a homolog of Xenopus XCAP-D2, colocalizes with phosphorylated histone H3 during the early stage of mitotic chromosome condensation. *Molecular and cellular biology* **20:** 6996–7006

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D & Jenuwein T (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes and Development* **18:** 1251–1262

Schultz DC, Ayyanathan K, Negorev D, Maul GG & Rauscher Iii FJ (2002) SETDB1 : a novel KAP - 1 - associated histone H3 , lysine 9 - specific methyltransferase that contributes to HP1 - mediated silencing of euchromatic genes by KRAB zinc - finger proteins. *Genes and Development* **16:** 919–932

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A & Cavalli G (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148:** 458–472

Sharp AJ, Spotswood HT, Robinson DO, Turner BM & Jacobs P a (2002) Molecular and cytogenetic analysis of the spreading of X inactivation in X;autosome translocations. *Human molecular genetics* **11:** 3145–3156

Shimura M, Toyoda Y, Iijima K, Kinomoto M, Tokunaga K, Yoda K, Yanagida M, Sata T & Ishizaka Y (2011) Epigenetic displacement of HP1 from heterochromatin by HIV-1 Vpr causes premature sister chromatid separation. *Journal of Cell Biology* **194:** 721–735

Shinkai Y & Tachibana M (2011) H3K9 methyltransferase G9a and the related molecule GLP. *Genes and Development* **25:** 781–788

Shintomi K & Hirano T (2011) The relative ratio of condensin I to II determines chromosome shapes. *Genes & development* **25:** 1464–1469

Shirai A, Kawaguchi T, Shimojo H, Muramatsu D, Ishida-Yonetani M, Nishimura Y, Kimura H, Nakayama JI & Shinkai Y (2017) Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *eLife* **6:**

Shogren-knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR & Peterson CL (2006) Histone H4-K16 Acetylation Control Chromatin Structure and Protein Interactions. *Science* **16:** 844–848

Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR & Peterson CL (2006) Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311:** 844–847

Simpson RT (1978) Structure of chromatin containing extensively acetylated H3 and H4. *Cell* **13:** 691–699

Smolle M & Workman JL (2013) Transcription-associated histone modifications and cryptic transcription. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1829:** 84–97

Sofueva S, Yaffe E, Chan W, Georgopoulou D, Rudan MV, Pollard SM, Schroth GP, Tanay A & Hadjur S (2013) Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO Journal* **32:** 3119–3129

Steen RL, Cubizolles F, Guellec K Le & Collas P (2000) Brief Report A Kinase – anchoring Protein ( AKAP ) 95 Recruits Human Chromosome-associated Protein ( hCAP ) -D2 / Eg7 for Chromosome Condensation in Mitotic Extract. *Cell* **149:** 531–536

Steffen PA & Ringrose L (2014) What are memories made of? How polycomb and trithorax proteins mediate epigenetic memory. *Nature Reviews Molecular Cell Biology* **15:** 340–356

Strunnikov a V, Hogan E & Koshland D (1995) SMC2, a Saccharomyces cerevisiae gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. *Genes & Development* **9:** 587–599

Sumner AT (1991) Scanning electron microscopy of mammalian chromosomes from prophase to telophase. *Chromosoma***:** 410–418

Sumner AT (1982) The nature and mechanisms of chromosome banding. *Cancer Genetics and Cytogenetics* **6:** 59–87

Tachibana M, Sugimoto K, Nozaki M, Ueda J, Ohta T, Ohki M, Fukuda M, Takeda N, Niida H, Kato H & Shinkai Y (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes and Development* **16:** 1779–1791

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K & Yamanaka S (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131:** 861–872

Takahashi K & Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126:** 663–676

Talasz H, Helliger W, Puschendorf B & Lindner H (1996) In vivo phosphorylation of histone H1 variants during the cell cycle. *Biochemistry* **35:** 1761–1767

Therizols P, Illingworth RS, Courilleau C, Boyle S, Wood AJ & Bickmore WA (2014) Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346:** 1238–1242

Th'ng HJP, Guo X-W, Swank RA, Crissman HA & Bradbury EM (1994) Inhibition of histone phosphorylation by staurosporine leads to chromosome decondensation. *J. Biol.Chem.* **269:** 9568–9573

Timms RT, Tchasovnikarova I a. & Lehner PJ (2016) Position-effect variegation revisited: HUSHing up heterochromatin in human cells. *BioEssays* **38:** 333–343

Tsumura A, Hayakawa T, Kumaki Y, Takebayashi SI, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, Nakayama JI & Okano M (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells* **11:** 805–814

Uemura T, Ohkura H, Adachi Y, Morino K, Shiozaki K & Yanagida M (1987) DNA topoisomerase II is required for condensation and separation of mitotic chromosomes in S. pombe. *Cell* **50:** 917–925

Vagnarelli P (2012) Mitotic chromosome condensation in vertebrates. *Experimental Cell Research* **318:** 1435–1441

Vagnarelli P, Hudson DF, Ribeiro SA, Trinkle-Mulcahy L, Spence JM, Lai F, Farr CJ, Lamond AI & Earnshaw WC (2006) Condensin and Repo-Man-PP1 co-operate in the regulation of chromosome architecture during mitosis. *Nature Cell Biology* **8:** 1133–1142

van Holde K & Zlatanova J (2007) Chromatin fiber structure: Where is the problem now? *Seminars in Cell and Developmental Biology* **18:** 651–658

van Kruijsbergen I, Hontelez S & Veenstra GJC (2015) Recruiting polycomb to chromatin. *The International Journal of Biochemistry & Cell Biology* **67:** 177–187

Verdaasdonk JS & Bloom K (2012) Centromeres: unique chromatin structures that drive chromosome segregation. *Nature Reviews Molecular Cell Biology* **114:** 65–84

Vos SM, Tretter EM, Schmidt BH & Berger JM (2011) All tangled up: How cells direct, manage and exploit topoisomerase function. *Nature Reviews Molecular Cell Biology* **12:** 827–841

Waizenegger IC, Hauf S, Meinke a & Peters JM (2000) Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103:** 399–410

Wang T, Xu C, Liu Y, Fan K, Li Z, Sun X, Ouyang H, Zhang X, Zhang J, Li Y, MacKenzie F, Min J & Tu X (2012) Crystal Structure of the Human SUV39H1 Chromodomain and Its Recognition of Histone H3K9me2/3. *PLoS ONE* **7**

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K & Peters JM (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451:** 796–801

Wigler M, Pellicer A, Silverstein S & Axel R (1978) Biochemical transfer of single-copy eucaryotic genes using total cellular DNA as donor. *Cell* **14:** 725–731

Wignall SM, Deehan R, Maresca TJ & Heald R (2003) The condensin complex is required for proper spindle assembly and chromosome segregation in *Xenopus* egg extracts. *The Journal of Cell Biology* **161:** 1041–1051

Wilkins BJ, Rall NA, Ostwal Y, Kruitwagen T, Hiragami-hamada K, Winkler M, Barral Y, Fischle W & Neumann H (2014) A Cascade of Histone Modifications Induces Chromatin Condensation in Mitosis. *Science* **343:** 77–81

Wood V, Gwilliam R, Rajandream M-A, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, et al (2002) The genome sequence of Schizosaccharomyces pombe. *Nature* **47:** 1215–20

Woodcock CLF, Frado LY & Rattner JB (1984) The Higher-order Structure of Chromatin : Evidence for a Helical Ribbon Arrangement Fibers from Interphase Nuclei and Chromosome. **99:**

Woodcock CL (1994) Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *Journal of Cell Biology* **125:** 11–19

Woodcock CL & Ghosh RP (2010) Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2:** a000596

Wu S, Hu Y-C, Liu H & Shi Y (2009) Loss of YY1 Impacts the Heterochromatic State and Meiotic Double-Strand Breaks during Mouse Spermatogenesis. *Molecular and Cellular Biology* **29:** 6245–6256

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C & Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442:** 86–90

Yamashita D, Shintomi K, Ono T, Gavvovidis I, Schindler D, Neitzel H, Trimborn M & Hirano T (2011) MCPH1 regulates chromosome condensation and shaping as a composite modulator of condensin II. *Journal of Cell Biology* **194:** 841–854

Yeap L-S, Hayashi K & Surani MA (2009) ERG-associated protein with SET domain (ESET)-Oct4 interaction regulates pluripotency and represses the trophectoderm lineage. *Epigenetics & Chromatin* **2:** 12

Yi Q, Chen Q, Liang C, Yan H, Zhang Z, Xiang X, Zhang M, Qi F, Zhou L & Wang F (2018) HP1 links centromeric heterochromatin to centromere cohesion in mammals. *EMBO reports***:** e45484

Yuan P, Han J, Guo J, Orlov YL, Huss M, Loh YH, Yaw LP, Robson P, Lim B & Ng HH (2009) Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes and Development* **23:** 2507–2520

Zatsepina O V., Polyakov VY & Chentsov YS (1983) Chromonema and chromomere: Structural units of mitotic and interphase chromosomes. *Chromosoma* **88:** 91–97

Zegerman P, Canas B, Pappin D & Kouzarides T (2002) Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *Journal of Biological Chemistry* **277:** 11621–11624

Zhang CL, McKinsey TA & Olson EN (2002) Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation. *Molecular and cellular biology* **22:** 7302–12

Zhang W, Bone JR, Edmondson DG, Turner BM & Roth SY (1998) Essential and redundant functions of histone acetylation revealed by mutation of target lysines and loss of the Gcn5p acetyltransferase. *EMBO Journal* **17:** 3155–3167

Zhao X, Ueba T, Christie BR, Barkho B, McConnell MJ, Nakashima K, Lein ES, Eadie BD, Willhoite AR, Muotri AR, Summers RG, Chun J, Lee K-F & Gage FH (2003) Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proceedings of the National Academy of Sciences* **100:** 6777–6782

Zhiteneva A, Bonfiglio JJ, Makarov A, Colby T, Vagnarelli P, Schirmer EC, Matic I, Earnshaw WC, Matic I, Mitotic EWC & Earnshaw WC (2017) Mitotic post-

translational modifications of histones promote chromatin compaction in vitro. *Open Biology*

Zhu B & Reinberg D (2011) Epigenetic inheritance: Uncontested? *Cell Research* **21:** 435–441

Zuin J, Dixon JR, Reijden MIJA Van Der, Ye Z & Kolovos P (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. **111:** 996–1001

Zullo JM, Demarco IA, Piqué-Regi R, Gaffney DJ, Epstein CB, Spooner CJ, Luperchio TR, Bernstein BE, Pritchard JK, Reddy KL & Singh H (2012) DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* **149:** 1474–1487